

Coming to terms with the human factor

The official report into the loss of the Mars Climate Orbiter raises many issues that need to be coolly scrutinized. High on the list are NASA's efforts to increase its cost-effectiveness by launching 'better, faster, cheaper' missions.

The expressions on the faces of managers from the US space agency NASA were grim last week, and their moods testy, when answering questions about the loss of the \$125 million Mars Climate Orbiter in September. And well they might be. An accident review board had just painted a disturbing picture of a top-notch technical organization gone awry (see page 221).

To judge by the board's conclusions, NASA's Jet Propulsion Laboratory (JPL), the undisputed world leader in the field of planetary exploration, has a serious management problem on its hands. But there is little time for soul-searching; the Mars Polar Lander is due to touch down on the red planet in two weeks, and work on the 2001, 2003 and 2005 Mars missions is under full, unrelenting way.

If JPL is too busy to step back and ask the question, maybe someone else should. Does 'better, faster, cheaper' really work? Or is it an empty slogan used to legitimize a chronically underfunded programme? It's difficult to prove either way. But a pattern is beginning to emerge from a string of recent mishaps. The demise of NASA's Earth-viewing Lewis spacecraft in 1997, and the near-demise of the US-European Solar and Heliospheric Observatory (SOHO) last year, were partly due to inadequate staffing — too few people doing too much, too fast.

At last week's press conference, Arthur Stephenson, who led the investigation into JPL's mishandling of the Mars Climate Orbiter, said that expecting the spacecraft navigation team to track three missions at once "may have been asking too much of them". In the case of Lewis and the Mars Climate Orbiter, an uncertain division of

responsibility between NASA and contractor personnel (at TRW and Lockheed Martin, respectively) also seems to have played a part.

All large projects have money and schedule pressures. But, along with being fabulously expensive (it costs \$125 million to send a camera and a spectrometer to Mars), space missions are unforgiving. A single, small mistake — which on Earth might be easily correctable — can lead to sudden, irretrievable loss.

This makes it imperative that NASA be brutally honest about how thinly its budget, which roughly translates to the number of people on the job, can be stretched. Bright, motivated scientists and engineers have a tendency to say "yes" to unrealistic challenges, then worry later how to pull them off. This may be happening now at JPL.

Not surprisingly, NASA science chief Edward Weiler and JPL director Edward Stone last week defended management's philosophy, which, after all, simply mirrors corporate culture in the late 1990s: do more with less, and downsize as much as possible. Congress and the White House have encouraged this approach, and should share fully in any blame, if blame is some day assigned.

All humans make mistakes, and experienced managers know to allow for this most unpredictable of factors. Stone acknowledged last week that his lab is still learning how to "do the programme differently" now that "low-cost programmes are the future". Such a change in culture should be encouraged. Meanwhile, NASA, Congress and the White House should stop relying on slogans and take a clear-eyed look at what the limits of 'better, faster, cheaper' really are. ■

Keeping the record straight on AIDS

Recent statements by South Africa's president on the hazards of AZT need a sympathetic but firm response.

Fate has not been kind to the new South Africa. Just when the country is struggling to overcome its political legacy and the economic consequences, it is being ravaged by the biggest medical scourge of modern times, AIDS. Yet the most effective weapons against the disease, such as the drug AZT (zidovudine) and its successor treatments, lie in the hands of Western-owned multinationals that many South Africans continue to see as having been at the root of many of their problems in the past.

In the circumstances, it is understandable that the government should choose to listen attentively to any criticism of such weapons, and — particularly when they carry a heavy price tag — to clutch at possible alternative solutions. Recent statements by South Africa's president, Thabo Mbeki, raising questions about the desirability of treatment using AZT, given its acknowledged toxicity, need to be evaluated in this context (see page 225). But taking such concerns seriously is one thing; endorsing policy statements that appear to be based on an incomplete understanding of the scientific principles involved, with potentially tragic consequences, is a different matter.

The criticisms currently being voiced in South Africa, indeed, over much of sub-Saharan Africa, where the disease has a stronger hold

than anywhere else in the world, are familiar. But they have also been carefully assessed, and in most cases firmly rejected, by those best placed to make such judgements, namely the world's leading scientific experts. For example, the response to the argument that HIV cannot be the cause of AIDS, as researchers are unable to explain precisely how it destroys the immune system, is that a complete understanding of the pathogenesis of a disease is no prerequisite to knowing its cause (a more detailed rebuttal of this and other myths about AIDS can be found on <http://www.niaid.nih.gov/factsheets/evidhiv.htm>).

Given the political context in which it currently operates, the South African government is not well placed to accept guidance on such issues, however valid, from private companies that could be seen as standing to profit commercially from it. But, hopefully, a domestic research community with no obvious axes to grind, and the weight of scientific experience behind its statements, can operate as an effective honest broker. The positions already taken by the Medicines Control Council and the Medical Research Council give optimism that sound reason will eventually prevail. Hopefully, the government will choose to listen to them rather than to others whose glib appeal masks an obscure agenda. ■





Journey to Mars
NASA learns the
lessons of
spacecraft failure
p221



Sell-off attacked
Privatization plans
for UK defence
research criticized
p223



In the dome
Human biology and
medicine among the
attractions
p224



Mbeki and AIDS
South Africa's
president stirs new
controversy
p225

Alliance of US labs plans to build map of cell signalling pathways

Munich

A prominent US cell biologist is seeking funds for a multi-laboratory, multidisciplinary initiative to map how molecules in a cell interact with each other in response to internal and external signals.

The project is being launched by Alfred Gilman, who won the 1994 Nobel prize in physiology or medicine for his work on the role of proteins in signal transduction. Called the Alliance for Cellular Signalling (AFCS), it would include systems engineers, biologists and informaticists, and is seen as a step towards the creation of a 'virtual cell'.

"The research community is doing a good job of describing signalling molecules and seeing how they interact, but we now need to put [the data] together in a large collaboration so that we can address the big question of how they all work together as a system," says Gilman. "Because of the availability of complete genome sequences, we can now approach the problem in way that is not biased towards particular proteins."

AFCS aims to identify the proteins that make up the signalling pathways in two types of mouse cell, the B lymphocyte (a cell of the immune system) and the myocyte (a heart cell). It also intends to assess the time-dependent flow of information through the systems in both normal and pathological states. Finally, it seeks to process this information and create theoretical models to describe cellular signalling.

Gilman, a professor of pharmacology at the University of Texas Southwestern Medical Center, has won support for the project from many key US cell-signalling researchers — nearly 40 top scientists have agreed to help to develop the project.

The AFCS would cost around US\$10 million a year to run, and would have two major components. First, it would create a network of around seven laboratories at academic sites across the United States, including the University of Texas Southwestern Medical Center at Dallas, the California Institute of Technology, and the University of California at San Diego and San Francisco.

These would generate information about signalling pathways in the two systems. Their

research efforts would be directed by two committees, whose chairs and members have already been chosen, as have the laboratory directors.

Much of the work would involve large-scale screening of protein-protein interactions, and would identify research leads that scientists elsewhere can follow up.

"We are not creating this big collaboration to dominate the field of cellular signalling," says Gilman. "Rather, we are trying to catalyse research in individual laboratories so that we can fill in the gaps in the signalling pathways."

He adds that alliance laboratories would not have privileged access to leads generated by the big screens they would undertake, as the information would be posted immediately on the Internet.

Scientists in the alliance laboratories have agreed to forgo most publishing and all commercial rights to their work, and pledge to



Small talk: adenylyl cyclase is one of the key molecules involved in cell signalling.

freely provide other researchers with their reagents and mutants. "It's a new way of doing business that requires collaborators

Scientists reel after Indian cyclone

New Delhi

India's worst cyclone this century, which ripped through the eastern state of Orissa on 29 October, flooded a nuclear accelerator lab, wiped out a botanical garden, and crushed the very radar system used to detect it.

The cyclone is also feared to have destroyed rice germplasm in the gene bank of the Central Rice Research Institute (CRRI) in Cuttack. One official estimates the damage at US\$8–10 million.

CRRI's rice gene bank had about 22,000 collections. Of these, the 5,000 or so varieties grown in experimental plots were washed away, says CRRI director Shanti Bhushan Lodh.

The viability of the remaining seeds, which are stored in a refrigerated module at 4 °C, is not known. But Rajendra Singh Paroda, head of the Indian Council of Agricultural Research, says: "We expect the seeds to be unharmed, as no water had entered the module."

Harder to replace will be the losses at the

Regional Plant Resources Centre, one of the world's largest botanical gardens, on the outskirts of the state capital, Bhubaneswar.

Its director, Premananda Dash, says the centre lost its entire germplasm of bamboo, palms, cacti and medicinal plants, and its collection of ferns, ornamental trees, shrubs and trees. "More than 20,000 tissue cultures have been swept away and equipment worth \$1 million lost," he says. "It will take eight to ten years to restore the centre."

The lost radar system was located at Paradip port. It was one of three warning radars on the east coast of India operated by the Indian Meteorological Department with the capability of detecting cyclone formations up to 400 km away.

At the Institute of Nuclear Physics in Bhubaneswar, the experimental hall of the 3-MeV 'pellatron' accelerator was flooded. The institute's library has disappeared, but Valangiman Ramamurthi, secretary to the Department of Science and Technology, says that the accelerator "is intact". **K. S. Jayaraman**

to act altruistically," says Gilman. At present, Gilman plans to keep alliance laboratories based in North America, and plans regular teleconferencing using the broad-band Internet 2 to link people and computers at the various centres. Collaborators cannot be spread across too many time zones, he says, and Internet 2 does not yet exist outside the United States.

But he hopes that the international community can become involved in the second part of AFCS's work, the creation of 'molecule pages' describing detailed properties of proteins involved in the signalling pathways.

These pages, which will be peer reviewed by an editorial board and regularly updated, will be created as a web database — or what Gilman calls a "virtual journal" — by 'alliance members'. These are currently being sought. AFCS's steering group sent out letters to the research community last month, and has so far received 110 replies, 10 per cent from non-US scientists.

AFCS plans to develop tools for data management and analysis that must also be flexible enough to accommodate a variety of systems-modelling techniques.

Gilman hopes to raise half the running



Gilman: creating a 'virtual journal'.

costs from a new grant programme of the National Institutes of Health's National Institute for General Medical Sciences, which is designed for large-scale interdisciplinary approaches to problems in biology. The grant proposal has reached the second round of assessment, and a decision is expected in spring.

Using a model similar to that of the SNP consortium (see *Nature* 398, 545; 1998), Gilman wants to raise the other half of the money from a consortium of pharmaceutical companies. He says these "will in any case benefit from the information about signalling systems, which are important drug targets". He is asking for an annual contribution of \$500,000 from each commercial participant.

Tony Pawson, a researcher at the Samuel Lunenfeld Research Institute at the Mount Sinai Hospital in Toronto, says the AFCS concept is timely because scientists are starting to create their own databases of the interactions

between cell-signalling molecules. But there is little coordination between them.

Pawson recently helped to launch a series of workshops under the auspices of the National Institutes of Health, the first of which was held in Washington last week. The aim is to help scientists talk to each other about where they can cooperate or merge their activities. He is talking to Gilman about how his own group's database, BIND (Biomolecular Interaction Network Database; <http://bioinfo.mshri.on.ca>) could cooperate, or even merge, with the AFCS databases.

Peter Seeburg, a director of the Max Planck Institute for Medical Research in Heidelberg, who studies mechanisms of action of membrane receptors, says that the initiative is "to be applauded". But because it is limited to two models, there is plenty of room for similar initiatives to be launched outside the project.

Axel Ullrich, director of the department for molecular biology at the Max Planck Institute for Biochemistry in Munich, says that the AFCS is "fantastic".

Full details of the new initiative can be found on <http://afcs.swmed.edu> **Allison Abbott**

Japan calls for partnership to develop semiconductors

Tokyo

A group of leading Japanese electronics companies has proposed a joint industry-government programme to develop next-generation semiconductor technology for the emerging system-on-a-chip (SoC) industry.

The proposal includes a plan to establish technical standards to enable intellectual-property components from different sources to be mixed and matched, which is necessary for the building of custom SoCs embedded within complex systems.

The move reflects a feeling within Japan's semiconductor industry — which once dominated more than half the world market — that the recent technological and industrial shift towards logic-based products may lead to a further decline in its world share, which has dropped to 25 per cent.

The proposal was made by Semiconductor Industry Research Institute Japan (SIRIJ), an organization funded by leading Japanese semiconductor manufacturers, including Fujitsu, Hitachi, Matsushita, Mitsubishi Electric, NEC, Oki Electric Industry, Rohm, Sanyo, Sharp, Sony and Toshiba, to promote industrial research into silicon semiconductor technology.

The programme would involve all the companies involved in SIRIJ, and is expected to obtain support from the Ministry of



Looking for integration: Japanese companies want to work with government to develop chips.

International Trade and Industry (MITI) to set up a core research centre in 2001.

The programme would be modelled on Project Alba in Scotland, which brought together Scottish Enterprise — a government economic development body — and universities and industry from both within and outside Scotland to promote the research and development (R&D) of SoC technology.

Although MITI says that the proposal is "not yet on the official agenda", industry sources say the plan is expected to be finalized by the end of this year and to be announced officially next spring.

"The Japanese semiconductor industry was slow to respond to paradigm changes in both technology and the market, as well as intensifying competition from Asian and US companies," says Tsuguo Makimoto, former director of Hitachi, who was

instrumental in drawing up the proposal.

Makimoto points out that Japan became too complacent about its dominance in memory-based technology, such as DRAM chips, in the 1980s. As a result, it "made very little effort in R&D for semiconductor technology for nearly a decade, while the rest of the world moved on".

As the global semiconductor industry makes a shift towards highly integrated devices for digital information products, the development of SoC technology, which enables complex systems to be put onto a single piece of silicon, is seen as crucial.

Over the next five years, advances in technology may allow up to 200 million transistors to be placed on a single chip, yet few companies have the intellectual property or technology to use this capacity.

"In addition to legal and technological challenges, the lack of adequate human resources is a serious problem for the Japanese semiconductor industry," says Makimoto.

He says that Japan needs an organization similar to the US Semiconductor Technology Council to oversee policies in semiconductor research, as well as a central institution to promote joint industry-university research and to design new curricula at universities. "In short, we have to forget our past glories and start everything from scratch," says Makimoto. **Asako Saegusa**

France could face 'scientific recession', warns council

Paris

A leading scientific advisory body to the French government, the Conseil Supérieur de la Recherche et de la Technologie (CSRT), last week strongly criticized France's research budget for 2000, warning that a failure to increase spending "could lead to a scientific recession in the near future".

The council says that the budget, drawn up by the science minister, Claude Allègre, represents an increase of only one per cent from the past year — less than the expected economic growth — to FF54.6 billion (US\$8.6 billion). "Public spending on research and development is falling behind other developed nations," the CSRT said in a statement.

The council also takes issue with the increase in the National Science Funds, a FF565 million fund established by Allègre last year for large projects or strategies delegated by the ministry. The council has demanded an evaluation of "this interventionist policy, for its impact on research and innovation as well as on the functioning of the research structures".

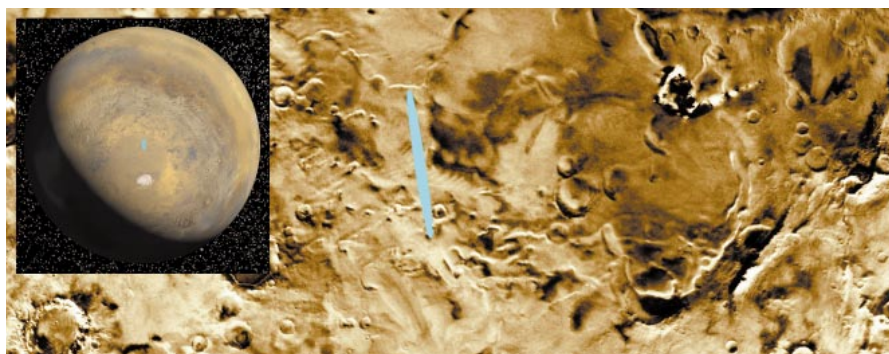
The CSRT emphasizes the need for these credits to be allocated in an open fashion. But it agrees with the ministry's plans to direct them towards laboratories and projects designed to stimulate technology and innovation.

Recruitment is another part of the budget over which the council expressed concern. France has an ageing scientific workforce, and it is estimated that 50 per cent of its current researchers will reach retirement age in the next ten years.

"The quality of French research in the future depends upon maintaining the quality of recruits in all of its corps," says the council. "The budget does not permit the anticipation of recruitment imposed by the wave of departures to retirement at the end of the next decade."

Further evidence of the tension between Allègre and his advisers comes from the resignation of Giorgio Parisi, a theoretical physicist and professor at La Sapienza University in Rome, from France's National Science Council, which was set up last year to advise the minister on large projects and research strategy. Parisi joins Nobel-prizewinning physicist Claude Cohen-Tannoudji and mathematician Yves Meyer, who resigned last month citing Allègre's refusal to discuss issues before making decisions.

Heather McCabe



Uncertain terrain: close-up views of planned Mars landing site show it to be rougher than expected.

Concern over Mars Lander as inquiry reports on Orbiter loss

Washington

The Mars Climate Orbiter was doomed by a combination of understaffing, poor communication, lack of training, and other management problems at the Jet Propulsion Laboratory (JPL) and its prime contractor, Lockheed Martin Astronautics. So concludes a report issued last week by a review board set up to investigate the loss of the \$125 million spacecraft in September.

Managers at the US space agency NASA are scrambling to add oversight staff and to take other corrective actions recommended by the board in preparation for a difficult landing on Mars in just two weeks.

The root cause of the loss of the Orbiter was Lockheed's failure to use metric units — as specified by NASA — in software used to calculate the spacecraft's trajectory (see *Nature* 401, 517; 1999). But the review board found a host of contributing factors. "There were many things in place that should have caught this error," says Arthur Stephenson, director of the agency's Marshall Space Flight Center in Alabama, who chaired the panel.

Spacecraft navigators and operators at JPL had noticed months earlier that their trajectory calculations were slightly off, but failed to appreciate the seriousness of the problem. The board described the navigation team as overworked and isolated from the rest of the JPL project staff.

The team had to handle three Mars missions at once, and were ignorant of key details about the Orbiter, wrongly believing it to have the same characteristics as the Mars Global Surveyor now in orbit around the planet. The navigation team's concerns about discrepancies in the trajectory data were not formally brought to the attention of senior managers.

The board also found that a last-minute trajectory correction manoeuvre that might have saved the spacecraft would not have been possible, because procedures to use it in an emergency were never put in place (see

Nature 401, 415; 1999). NASA plans to develop those procedures for next month's arrival of the Mars Polar Lander.

Lack of communication and unclear lines of authority plagued the project, and still do, according to the report. "A recurring theme in the board's deliberations was 'who's in charge?'"

Defensiveness over the loss of the Climate Orbiter is further hindering cooperation between different elements of the project, wrote the panel. JPL director Edward Stone said last week that the lab will add about 30 senior engineers to the project, and has created an advisory group of six senior spacecraft navigators in an all-out effort to ensure that no problems are being overlooked for the 3 December Mars landing.

While the board's inquiry focused on JPL, a separate review of the lab's relations with its contractors, released by the NASA Inspector General in September, highlights problems at Denver-based Lockheed Martin, which is building NASA's fleet of Mars spacecraft.

In the era of 'better, faster, cheaper' projects, the agency is relying more than ever on quality control at companies such as Lockheed. But JPL "does not have adequate policy for monitoring subcontractor performance", according to the Inspector General's report.

Lockheed understaffed the Mars project in its design phase, which may have increased the likelihood of problems being introduced early. The contractor then had people working 70-hour weeks at the end of the project to meet the spacecraft delivery date. "This could have an adverse effect on employee morale and increases the potential for human errors," said the Inspector General.

Stone and NASA science chief Edward Weiler staunchly defended the agency's 'better, faster, cheaper' philosophy last week. They blamed the loss of Mars Climate Orbiter on a failure to follow rules and procedures already in place.

Louis Friedman, executive director of the

Planetary Society and a veteran observer of JPL, also rejects the idea that NASA's Mars programme is underfunded, or that more money will necessarily fix the problem. "You can do things wrong at any price," he says.

But the programme is clearly showing signs of strain. Weiler's office quietly announced its decision last week to postpone indefinitely a planned Mars airplane demonstration flight in 2003, even though industry teams were invited to bid for the project a few weeks ago.

The next landing will be more difficult than the 1997 Pathfinder landing, yet is being done at lower cost. The Mars Polar Lander will use legs rather than airbags to touch down on the planet's surface, and a large boulder in the wrong place could spell disaster.

Recent high-resolution photos of the targeted landing area show that a small fraction of the site appears rougher than expected. Science planners were not worried enough to switch to a back-up site last month but, as with any planetary landing, they will be crossing their fingers at the critical moment.

The Polar Lander will also use a new kind of descent engine with pulsed rocket jets, which raised concern among the members of the accident review board. "This type of powered descent maneuver has always been considered to be very difficult and stressing for a planetary exploration soft landing," they wrote.

But planetary scientist David Paige of the University of California, Los Angeles, a principal investigator for the Lander mission, says he is confident that the JPL team has looked carefully at potential problems. **Tony Reichhardt**

Professors use web to catch students who plagiarize ...

Berkeley

A computer-based service to detect student plagiarism is being used by a growing number of university teachers across the United States, and may soon be tested in Britain. The service was developed by a doctoral candidate at the University of California, Berkeley.

Called Plagiarism.org, the service was launched last spring. It allows academics, journal editors — and even students — to rapidly compare articles against thousands of papers available through the Internet. The Internet has itself made plagiarism far easier through simple cuts and pastes.

One German researcher says he has already found the programme so effective that he plans to scrutinize all manuscripts submitted to his recently launched publication, the *Journal of Medical Internet Research* (see below).

Plagiarism.org was set up by neurophysiology doctoral student John Barrie. According to Doug Zuidema, head of the Office of Student Conduct at the University of California at Berkeley, the university is negotiating a contract with the company.

"We hope to negotiate an agreement for all faculty and students on the campus to have access to the service," says Zuidema. Plagiarism.org is already negotiating for a pilot study to be carried out at selected British universities through the UK Higher



Education Funding Council.

The program originated from Barrie's experience in grading neurobiology papers, when he became suspicious that students were plagiarizing material. In some cases they were taking material from online companies that sell articles over the Internet via websites such as Schoolsucks.com and Cheater.com.

After working as a teaching assistant with psychopharmacologist David Presti, Barrie created the service and enlisted the help of several Berkeley graduates as partners. Last spring, he used the system to check the papers of about 320 juniors and seniors in Presti's upper-division neurobiology class.

When a paper or article is run through the Plagiarism.org vetting system, a printout is generated on which apparently plagiarized material is highlighted. This is done by linking apparently cribbed sections to their sources, typically published articles.

At the beginning of the semester, Presti had told the students that their papers would be checked for plagiarism. Afterwards, Barrie examined the papers and found that 15 per cent of the students had plagiarized material.

One student author from Presti's class, for instance, appeared to have based virtually his whole article on sections of published work lifted from six web addresses.

"Clearly this is a serious problem; we have to do more analysis of it," says Paul Licht, dean of Berkeley's College of Biological Sciences. "I'm not sure whether I'm more discouraged that they committed plagiarism, or that they continued to do so after they were warned."

Presti and Berkeley officials are reviewing possible action against the students. They could be given a failing grade for either the paper or the class, or, more seriously, be charged with violating the code of student conduct.

Rex Dalton

... and author gets similar paper retracted

San Diego

Gunter Eysenbach of the department of clinical social medicine at the University of Heidelberg, and editor of the *Journal of Medical Internet Research*, plans to publish a report on the uncovering of apparent plagiarism by three physicians from the Royal Infirmary in Edinburgh, Scotland.

After their article, entitled 'The quality of surgical information on the Internet', was published in the August issue of the *Journal of the Royal College of Surgeons of Edinburgh*, the authors sent out e-mail alerts to those interested in online medicine.

Eysenbach saw the alert, read the paper and recognized phrases in it as appearing to come from an

article that he had written for the *British Medical Journal* in October 1998. "I am not a native English speaker, so it takes hard work to write good English sentences," says Eysenbach. "I recognized the anguish of my work."

Eysenbach also discovered that other material from his journal's website had been used without attribution in the article. Aware of Plagiarism.org (see above), he registered with the US service's website and submitted the article by the Edinburgh physicians, without disclosing what he already knew.

According to Eysenbach, the analysis detected the apparent plagiarism, as well as the improper use of other published material. "About 50 per cent of the

article is affected," he claims, adding that the senior author — Christopher Oliver, a trauma surgeon at the Royal Infirmary in Edinburgh — initially offered several explanations, from denial to forgetting references.

Ultimately, he says, Oliver agreed to retract the article and apologize. Oliver says that "it was an accident; there was no intent to plagiarize". He added: "I have better things to do than plagiarize his work. It was an omission on my part not to give references."

Although he acknowledges his retraction and apology, Oliver says that "If you ran [this system] on every article [in the medical literature] that comes out, you would find this happening all over the place."

R.D.

Privatization of UK weapons labs under fire

London

Proposals for the partial privatization of one of Europe's largest research organizations, the UK Defence Evaluation and Research Agency (DERA), have been criticized by a cross-party committee of the British Parliament.

There is concern that the move could undermine collaboration in defence research with the United States, whose government will only deal directly with other governments, and not with private organizations. US officials are said to be unhappy about the proposed privatization, pointing out that certain areas of development of military technologies would become out of bounds to DERA institutions.

DERA exists primarily to support the UK Ministry of Defence (MoD), for example by running most of its non-nuclear research, technology and test and evaluation establishments. But the agency has increasingly been encouraged by the government to take on commercial work. The Treasury now wants DERA to become essentially a private organization, requiring it to become more financially competitive.

But the defence select committee of the House of Commons says it has found several areas in which the MoD's proposals for a public-private partnership (PPP) are "flawed or where significant issues remain to be resolved". These include the difficulty of protecting collaboration with the United States.

"The collaboration between Britain and the United States is the closest in terms of defence technology. It goes back 40 years," says Philip Gummett, a research policy analyst at the University of Manchester.

Some claim that the proposals would also mean that officials from DERA would no longer be able to take part in international governmental organizations and committees, such as those of the North Atlantic Treaty Organization. This would raise the issue of who would provide scientific and technical input on behalf of the UK government at such meetings.

The Institute of Professionals, Managers and Specialists (IPMS), the labour union that represents most DERA employees, submitted a lengthy statement to the select committee arguing against privatization. The union's concerns are reflected in the committee's report. As well as DERA's relationship with the United States, the extent to which the MoD would lose its 'intelligent customer' facility was criticized.

At present, DERA provides most of the scientific and technical expertise used by the MoD to assess any equipment programme. Charles Harvey of IPMS says privatization could leave the MoD "bereft of expertise". "In



Misfire? A rifle system designed by DERA, whose US links could be under threat.

the taxpayer's interest, it is essential that there is intelligent customer capability in the MoD. Hiving off its expertise would not be a good idea. This is one of the features of the report that we particularly applaud."

Another stumbling block is the political sensitivity of some of the research carried out by DERA — in particular by its Chemical and Biological Defence Sector at Porton Down — and whether this should be entrusted to a private company.

Widespread opposition to the Treasury-led proposals has already led the defence

secretary, Geoffrey Hoon, to announce a second postponement of a decision on the matter last month. But continued pressure on UK departmental spending mean that the MoD will not be shelving PPP plans.

Hoon said the MoD would widen the scope of discussions to "address the concerns raised" in the report. He added: "We are committed to DERA PPP and are confident that we can achieve a solution which will strengthen DERA's ability to provide world-class scientific research while preserving our defence interests and maintaining our collaborative relationships."

Options being considered include keeping a larger part of DERA within the MoD. Spending on defence research has fallen by 40 per cent over the past six years, and the committee says that the ministry remains committed to a continual downward course for research spending.

IPMS officials welcomed the conclusions of the select committee, saying that they echo the union's concerns. The MoD's two postponements of its plans for PPP shows its difficulty in implementing them, says Harvey. "The proposals should be redrafted at the least, but we would hope they are withdrawn." The way forward, he says, is to develop the Defence Diversification Agency and to diversify DERA into more non-MoD related work.

Natasha Loder

Jobs boost for Spanish science

Barcelona

The Spanish government last week committed itself to creating 2,000 new posts in the public research sector over the next four years, and to raise the overall number of researchers in the workforce from 3.3 to 4 per 1,000 employees.

Two types of job will be created: newly qualified PhDs will be offered research contracts of up to five years in publicly funded research centres, in priority areas identified under the national plan, and a number of more senior researchers with at least ten years research experience will be given five-year contracts that will be renewable for a further five years.

The commitment is part of a four-year National Plan on Research, Development and Technological Innovation that will come into effect on 1 January. Mariano Rajoy, the minister of education and culture, said the plan was one of the Conservative government's "most important issues".

Central to the strategy is coordination and integration between ministries responsible for research and development

(R&D), and improved evaluation of government-funded research projects.

The Interministerial Commission of Science and Technology, which is responsible for coordination, will be strengthened, as will the public bodies responsible for assessing research projects, including the National Agency of Evaluation and Prospective and the Centre for Technological and Industrial Development.

Under the plan, public spending on R&D is expected to grow by at least 6 to 8 per cent a year until it reaches 1.3 per cent of the gross national product by the end of 2003 (see *Nature* 400, 393; 1999).

Next year's government research budget will be Pts508 billion (US\$ 3.1 billion), an increase of Pts48 billion over this year. Including private sector and regional expenditure and funds received from the European Commission, research spending will total Pts1,500 billion.

The plan includes measures designed to support applied research and technological innovation in Spanish companies. Josep Piqué, the minister of industry and energy,

► says that the government wants the proportion of companies engaged in high-technology projects to grow from 12 to 20 per cent by the end of 2003.

As part of this trend, the government will provide financial support for 500 young postdocs to be recruited by Spanish companies, as well as 1,000 scientists in technology centres and small- and medium-sized companies.

There will be tax incentives for companies involved in R&D, including a 30 per cent increase in general tax deductions and a further 10 per cent deduction for expenses related to research personnel.

Piqué points out that Spain's position on technological innovation is "not good" compared with other developed countries, and that a "great effort" will be needed, mainly from the private sector, to reach a similar level to its competitors.

He says that Spain must move beyond a 'subsidy culture', and that even though the country is ranked eleventh in the world in terms of its research output, its technological exports are "very small" compared with its imports.

At a recent meeting in Valencia with 2,000 industrial managers, Spain's prime minister, José Maria Aznar, said his government wanted investment by the private sector in R&D to grow by more than 10 per cent a year, to reach 65 per cent of the country's total R&D budget by the end of 2003.

Many scientists have welcomed the new posts and the budget increase. But they warn that the plan will not solve the bias towards local candidates in selection procedures.

Xavier Bosch

Opposition pledges to raise New Zealand science budget

Sydney

New Zealand's opposition Labour Party has promised to increase the funding for basic research by NZ\$77 million (US\$40 million) over the next three years, financed primarily from increased taxes on high-income earners, if it wins the general election next week (27 November).

The party is also promising increased taxes to raise an extra NZ\$750 million over three years, most of which would be used to ease student debt, stop fee increases and encourage students to take up science and technology.

The ruling National Party, which recently attempted to take the initiative with measures — but no new money — intended to boost the nation's research and development (see *Nature* 401, 106; 1999), appears to have lost some momentum on the issue.

Many scientists were disappointed that the prime minister, Jenny Shipley, gave only passing mention to research in her campaign launch, strongly defending the National Party's belief in market forces. Indications that the party would reverse its opposition to tax breaks to improve the poor research performance by industry were scotched last week by the treasurer, Bill English.

But the government-owned Association of Crown Research Institutes, through their association, are strongly supportive of the earlier package, and have declared that "it puts research centre stage in New Zealand's development".

Both Labour and the National Party have pronounced that the 'knowledge economy' is the country's route to economic salvation. But science is a less popular cause than universities, where staff and students are protesting against government policies.

These policies, which have led to rising fees and student debt, are widely cited as being behind the exodus of graduates and experienced researchers alike.

This week, New Zealand-born scientists working in the United States claimed the government's package of new scholarships would be better spent on improving research conditions in New Zealand.

Amanda Peet, a physicist at the University of California, Santa Barbara, says that higher education in New Zealand is "starved of support".

Labour's science spokesman, Mark Peck, is promising to establish New Zealand's first external body to advise the prime minister on science and technology — a significant weakening of the ministries' grip on policy.

Many are critical of the government. Mark Grimes, a member of the council of the New Zealand Association of Scientists, accuses the government of "a plot to cover up their abysmal record in funding science and education by quoting cumulative increases".

But others are sceptical of Labour's capacity to improve the situation, given that it started the problem with its cuts. The association's president, Janet Grieve, says it is not endorsing any party.

Peter Pockley

Body takes centre stage in UK's Millennium Dome



London

Some of the main attractions inside the Millennium Dome — the centrepiece of the British government's year 2000 celebrations — were unveiled last week. The Body Zone (pictured) is designed to give visitors an insight into their bodies and the discoveries, technologies and new ideas that will affect health and well-being in the future.

Visitors will walk through the giant figure and move on to exhibits on the future of drug design, robots at work in the operating theatre of the future, and an iris scanning display. The Human Genome Project will be featured in a wall of telephone directories representing the quantity of information in human DNA.

Located at Greenwich in London, the dome is expected to receive up to 12 million visitors next year.

Natasha Loder

AZT critics 'swayed South African president'

Cape Town

South Africa's Minister of Health, Manto Tshabalala-Msimang, last week defended claims by the country's president, Thabo Mbeki, that the anti-retroviral drug AZT (zidovudine) may be too toxic to give to patients with HIV (see *Nature* 402, 3; 1999).

Her statement came in spite of an interim report from the country's statutory body for drug control, the Medicines Control Council (MCC), compiled by the council on its own initiative after Mbeki's claims. This concludes that, although AZT has some negative side effects, they are outweighed by its potential advantages.

But there is also evidence that the president may have been influenced by a Pietermaritzburg-based lawyer, Anthony Brink, who appears to hold views similar to those of the US molecular biologist Peter Duesberg — described by Brink as “a paragon of scientific integrity in relation to conventional orthodoxy”.

Tshabalala-Msimang said that the government would not make AZT routinely available at state hospitals until it receives the MCC's full report, due in two months. But she added that the drug would not at this stage be withdrawn, and that those receiving it should continue to be treated, pending the final report.

Peter Moore, research director for sub-Saharan Africa for Glaxo-Wellcome, met with Tshabalala-Msimang last week. He described the meeting as “cordial”, but says the minister was unable to identify any specific safety concerns, or any specific side effects that she was concerned about.

According to Mbeki's spokesman, the president got his information on the drug from the Internet. But Brink, a barrister specializing in medical litigation who has been campaigning against the use of AZT, claims that he was the source, having provided information to Tshabalala-Msimang. He also says that Mbeki's office contacted him for more information after the president's statement.

Brink wrote a review article on the alleged dangers of AZT after a debate earlier this year in *The Citizen*, a Johannesburg newspaper — for which he had written an article entitled “AZT: a medicine from hell”. A rebuttal was written by Des Martin, president of the Southern African HIV/AIDS Clinicians Society.

Brink's review article lists papers in a number of medical journals that he claims show the dangers associated with the drug. He says that the “scientific accuracy of his article is impeccable”, and has been checked by researchers in Australia and Florida.

Brink denies that Duesberg, an outspoken proponent of the view that AZT is unac-



Hot seat: Mbeki, seen here at a national AIDS broadcast last year, is under pressure to act.

ceptably dangerous, is the progenitor of his ideas. But he admits that Duesberg has congratulated him on his article, and encouraged him to submit it to a scientific journal. The two share the view that AIDS is not related to any specific virus.

But immunologist Malegapuru Makgoba, president of the Medical Research Council (MRC), describes the grounds of Brink's argument as “nonsensical”. He adds: “I've

read nothing in the scientific or medical literature that indicates that AZT should not be provided to people.”

Makgoba stresses the need for government policies to deal with three issues which, he says, have become confused in the minds of the South African public: the AIDS epidemic, the high incidence of rape in the country and the provision of AZT by the state.

Mbeki's statement is widely seen as an attempt to justify his government's failure to develop such policies. Salim Abdool Karim, head of AIDS research at the MRC, has been reported as saying that “if the president doesn't want to provide AZT, he should find an excuse based on fact”.

Meanwhile, the government has laid five charges of misconduct against Costa Gazi, head of public health at Celia Makiwane Hospital in East London, for allegedly contravening the Public Service Act. The charges arise from an article in a local newspaper earlier this year in which Gazi — who is health secretary of the opposition Pan Africanist Congress — was quoted as criticizing the government for not providing AZT to pregnant women with HIV (see *Nature* 396, 504; 1998).

Mike Cherry

Russian peace researcher detained

London

The Russian Federal Security Service (FSB) has detained a Russian researcher, Igor Sutyagin, who works on arms control, disarmament and security problems, and interrogated two of his colleagues.

Sutyagin, head of the military-technical research section at the US-Canada Institute of the Russian Academy of Sciences, was detained on 27 October by the FSB, the successor to the Soviet KGB. His apartment in Obninsk, 100 kilometres south of Moscow, was searched and officers removed computers and papers.

Later that day FSB officers searched the Moscow apartment of Joshua Handler, a PhD student at Princeton University working on a dissertation on nuclear arms control in the 1990s between Russia and the United States. He is working under Frank von Hippel, vice-president of the Federation of American Scientists.

Officers removed Handler's computer, notebooks, address books, papers, photographs and maps, although they left his passport, money and air ticket. Eight officers spent until late evening searching Handler's apartment and questioning him about his work.

Handler has been a colleague of

Sutyagin's, and the US-Canada Institute, for many years. He has now returned to the United States for what he calls a temporary “vacation”.

Handler has worked on arms control for 15 years. He says: “My work has been directed at understanding and improving US-Russian relations in the hope of encouraging greater steps towards nuclear disarmament.” He adds that “it has never, and was never intended to, harm any country's security”.

Handler was until recently a researcher for Greenpeace International. He worked on its nuclear disarmament campaign in the 1980s and 1990s, writing many articles and papers about nuclear weapons and the attendant military and environmental problems.

“I've always published [my] research results openly and sought to disseminate [them] as widely as possible,” he says, adding that it is unclear why Sutyagin has been detained.

In response to Sutyagin's detention, Sergey Rogov, director of the US-Canada Institute, has issued a statement saying: “Like other institute employees, Sutyagin does not have access to information constituting state secrets.”

Carl Levitin

Publishers agree on deal to link journals on the web

Paris Plans by commercial scientific publishers to link references in the articles they publish to the source papers on the websites of their respective publications moved a step forward this week, with the announcement of a formal agreement between 12 major publishing houses, including *Nature*.

The publishers said that as many as three million articles across thousands of journals would be linked by early next year (see *Nature* **402**, 115; 1999). The agreement will be managed by an elected board, which will cooperate with the International Digital Object Identifier Foundation (<http://www.doi.org/>). This is trying to develop a unique code for every web page that would stay with it throughout its life, irrespective of its physical location.

Floyd Bloom, editor-in-chief of *Science*, describes the venture as “a major step forward in gaining access to the recent body of literature, greatly surpassing any other private or government-sponsored resource”. The publishers behind the agreement are: Academic Press; American Association for the Advancement of Science; American Institute of Physics; Association for

Computing Machinery; Blackwell Science; Elsevier Science; The Institute of Electrical and Electronics Engineers; Kluwer Academic Publishers; Nature; Oxford University Press; Springer; and John Wiley & Sons.

US pressure groups seek moratorium on GM foods...

Washington US pressure groups opposed to genetically modified (GM) crops have released a Pacific Declaration, calling for a moratorium on the release of new GM foods. It is timed to coincide with the World Trade Organization's meeting at Seattle later this month.

The declaration, whose initial supporters include Friends of the Earth, the Rural Advancement Foundation International and Jeremy Rifkin's Foundation on Economic Trends, calls for labelling of GM foods as well as stricter regulation of the crops. The pressure groups are launching a large-scale advertising campaign attacking agricultural biotechnology in US newspapers.

...but consumers 'get a taste for biotech'

London Two-thirds of US consumers support the use of biotechnology to produce foods and have confidence in the policy of the Food and Drug Administration (FDA) for labelling

biotech foods, according to a survey of 1,000 adults carried out for the industry-sponsored International Food Information Council.

The survey found that seven out of ten Americans support FDA labelling policy, which requires that foods produced through biotechnology have special labels only if the food has been significantly changed. Some 73 per cent of respondents had heard “something about biotechnology”, and 63 per cent expected to benefit from it over the next five years.

Panel seeks way ahead for troubled US laser facility

Washington A panel chaired by John McTague, a former senior executive at the Ford motor company, has started a speedy investigation of options available to complete the US National Ignition Facility. This troubled \$1.2 billion laser experiment is being built at the Lawrence Livermore National Laboratory in California.

The inquiry started on 15 November by hearing testimony from Bruce Tarter, director of the laboratory, and others about reported delays and projected cost overruns on the project, one of the largest scientific facilities under construction in the United States. The panel is expected to report next month to Bill Richardson, the energy

secretary, on ways forward. These may include initial construction of 96 lasers, instead of the 192 in the original design.

Strike removes rector of Mexican university

Washington Francisco Barnés de Castro, rector of the National Autonomous University of Mexico, has resigned after seven months of a bitter strike and occupation by students that paralysed teaching and research at Mexico's largest university.

The strike was a response to an attempt by de Castro to increase annual tuition fees from a nominal sum to about US\$140, and his resignation will be seen as a victory by the students. The university has already withdrawn the fee increase, but students have extended their demands to include easier access to the university.

Royal Society demands urgent action on salaries

London Britain's Royal Society has responded to an independent inquiry into higher education by calling on the government to improve academic pay and conditions (see *Nature* 399, 719; 1999). A report issued last week states that uncompetitive remuneration packages are making it difficult for universities and col-

leges to recruit and retain academic staff of the highest calibre. Without urgent action, says the society, Britain's world-class science base will suffer "irreparable damage".

The society calls on the government to increase minimum academic salaries and to allow local flexibility in salary setting, and adds: "It should be possible to offer longer contracts, on higher salaries."

German initiative aims for younger professors

Munich The age at which young scientists in Germany can become tenured professors could be reduced by introducing career structures similar to those at US universities, according to the science ministers of Germany's 16 *Länder* (states). At the end of a recent meeting, the ministers said the target age should be 34, and the qualification should consist of a two-year postdoctoral period and a four-year period as assistant professor.

In contrast to the traditional German system — the widely criticized *Habilitation* — the research of assistant professors should be independent from that of established professors, the ministers said. The proposals have been sent to a committee set up by federal research minister Edelgard Bulmahn, which is working on how to change academic employment rules.

CNRS head to lead panel of research council chiefs

Paris Catherine Bréchignac, director-general of France's Centre National de la Recherche Scientifique, has been nominated to chair the group of European heads of research councils (Eurohorcs), which meets every six months to discuss science policy issues. Bréchignac replaces Richard Brook of Britain's Engineering and Physical Sciences Research Council.

Supercomputer centre opened in Russia

Moscow An Interdepartmental Supercomputer Centre was opened in the new Presidium building of the Russian Academy of Sciences in Moscow last week. It was set up by the academy, the Ministry of Science and Technology, the Ministry of Education and the Russian Foundation for Basic Research.

The supercomputer centre, the biggest in Russia, is equipped with a Hewlett-Packard V2250 system and a Russian-made MVS-1000 system, produced by the Kvant scientific institute. It will be accessible to 760 registered users in 40 regions and 32 universities. "This is the result of basic science of the highest standard," said Vladimir Putin, the prime minister, who opened the centre.

Genetically modified foods face rigorous safety evaluation

Sir—The Commentary by Erik Millstone *et al.*¹ on the role of substantial equivalence in the safety evaluation of genetically modified (GM) food draws attention to important issues, but it gives an inaccurate and misleading account of the work of regulatory committees and the current role of substantial equivalence in safety evaluation.

The authors imply that regulatory committees depend on substantial equivalence as the sole basis for the safety evaluation of GM foods. They suggest that only chemical tests are used, and that biological, toxicological and immunological methods are ignored. This seriously misrepresents the work of independent expert committees that give careful and wide-ranging consideration to the safety of GM foods on a case-by-case basis, and do not rely solely on substantial equivalence.

From the standpoint of safety evaluation, substantial equivalence is a useful tool to address a major limitation in traditional toxicology approaches to whole GM foods. The feeding of excess quantities of individual chemical components to experimental animals can easily be distinguished from overall nutrition. This approach can be used to set acceptable daily intakes for non-nutrient chemical components of the diet. In contrast, whole GM foods are complex, contribute to nutrition and are limited in the quantity that can be consumed. So the interpretation of data from animal feeding experiments is far from simple.

The safety evaluation of GM food is far more rigorous than is applied to its conventional counterpart, and aims to establish that the accepted safety of the conventional counterpart has not been compromised. In Europe, the safety evaluation of GM foods is covered by European Commission (EC) regulation 258/97. The work of the UK Advisory Committee on Novel Foods and Processes (ACNFP) now falls within its framework². According to the EC, "substantial equivalence is not a safety or nutritional assessment in itself, but an approach to compare a potential new food with its conventional counterpart"³.

In most instances GM technology is applied to a crop plant to introduce a new trait, so the GM plant cannot be substantially equivalent to its conventional parent. An example that illustrates the fact that much more than substantial equivalence is involved is the safety evaluation of GM soya⁴. Safety evaluation focused on four issues: intentional changes; unintentional changes; stability; and gene transfer. GM soya contains a single

introduced bacterial gene that confers tolerance to the herbicide glyphosate. This gene produces a glyphosate-insensitive homologue of the natural plant enzyme, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), that is involved in aromatic amino-acid biosynthesis. Detailed molecular analysis authenticated the genetic modification and demonstrated that it was stable through several generations of conventional plant breeding. An acute toxicity study in mice showed that the introduced EPSPS protein was non-toxic, and it was demonstrated that this protein is rapidly degraded under conditions encountered following consumption. Furthermore, the conventional processing of soya was shown to destroy the protein. So EPSPS, which in any event is inactivated in marketed GM food, is safe.

The possibility that unintended changes had taken place was evaluated by comparing compositional data and nutritional studies for GM and conventional soya. It is known that unprocessed soya beans can cause food allergy, and the levels of known allergenic proteins were unchanged in GM soya. The possibility of gene transfer from GM soya was eliminated by DNA degradation during processing. On the basis of this evaluation, ACNFP concluded that GM soya was as safe as its conventional counterpart.

Substantial equivalence has recently been re-evaluated by the Organization for Economic Cooperation and Development⁵. A useful aspect of this evaluation is the recommendation of a consensus on appropriate components for compositional analysis so as to standardize safety evaluation. Key nutrients and known toxins, antinutrients and allergens would be included in this consensus.

The expression of previously unknown toxins, antinutrients or allergens in GM foods so as to cause previously unrecognized harm is unlikely, given that conventional foods have been subject to massive changes in genetic make-up by established plant-breeding methods. The use of chemical fingerprinting, messenger RNA analysis, DNA arrays and proteomics to investigate unintended effects are recognized possibilities for the future, but the practical value of these techniques has not been established. Changes to gene expression and to the levels of individual proteins and metabolites are a normal feature of plant development and response to environmental change. Any such detailed holistic analysis needs to be considered in the context of a dynamic situation in which flux in gene expression is the norm.

M. J. Gasson

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK

1. Millstone, E., Brunner, E. & Mayer, S. *Nature* **401**, 525–526 (1999).
2. ACNFP Annual Report 1997 3–6 (1997).

3. *Official Journal of the European Communities* **L253**, 16/09/1997, 1–36 (1997). (http://europa.eu.int/eur-lex/en/lif/dat/1997/en_397X0618)
4. ACNFP Annual Report 1994 59–65 (1994).
5. <http://www.oecd.org/ehs/food>

Bureaucracy blights Japan's safety record

Sir—Your editorial, "Perils of inadequacies in safety regulation", hits on a universal problem in Japan (*Nature* **401**, 513; 1999). Besides the nuclear accident and blood products scandal you refer to, there are numerous other examples that testify to the absence of proper management of science and technology in Japan. Large quantities of dioxins are still emitted from refuse incinerators despite global concerns, for example.

Why is Japan unable to control these serious problems? The main reason lies in the machinery of national government, in which the system of authority is administered by bureaucrats. Government offices jettison their veteran officials into high positions in public corporations and related industries to maintain their pyramidal stratification based on a system of seniority. According to the prime minister's office, of 6,843 public corporations under the jurisdiction of the national government, 2,470 had 6,903 ex-bureaucrats as directors in 1997. But this is only the tip of the iceberg. There are many other corporations under the jurisdiction of local government, and numerous private enterprises, that accept retired bureaucrats as executives. The nuclear and pharmaceutical industries are no exception. This practice has meant that respect for irresponsible figures in authority has been cultivated and protected.

The bureaucrats have excellent administrative ability, and industries find it valuable to make connections with them. But these bureaucrats lack expertise in management of unexpected disasters, such as the Tokai nuclear accident. To make matters worse, they lack the expertise needed to regulate the industries for which they are responsible.

Unless this problem is resolved, it will be impossible to establish in Japan effective regulatory bodies, similar to the US Food and Drug Administration. The bureaucrats have a strong hold on major industries. They know that, if new regulatory bodies are established, their grip on industry will loosen irreparably. And industrialists feel they may not survive unless favoured by authority — many companies have made lucrative profits under the status quo.

The first step to resolve this problem may be international acknowledgement of the lack of expertise in safety regulation among Japanese bureaucrats. Responsibility for the management of science and technology

should not be limited to one country. We have more powerful science and technology than ever before, and this does not allow any margin for errors. The Tokai accident and other scandals sound a warning that Japan, which should have high levels of control of science and technology, has made light of safety regulation.

Kazuo Inoue

Towa Clinic, 468 Showa, Towa, Hata,
Kochi 786-0511, Japan

Challenge for global e-journal project

Sir—The proposed repository for research reports, PubMed Central¹, is rapidly decomposing itself. As I cautioned in *Nature*², the US National Institutes of Health (NIH), which developed the proposal, has failed to confront three key issues: funding, lack of expertise, and consensus building. The latest silliness involves an alliance between PubMed Central and the Community of Science, to fund the peer-review function of the repository. This is yet another bleak development in the brief history of this initiative.

NIH director Harold Varmus proposed this project in a particular setting. Workers from many disciplines, including library and information sciences, computer science, clinical medicine, and the biological sciences, had been struggling for several years to develop the tools and procedures needed to produce an online archive of bioscience research reports. When Varmus and the NIH were initially confronted with an avalanche of negative comments, they quickly attempted to revise the proposal. But each succeeding incarnation of the PubMed Central scheme was ever more insipid. Let us remember the philosophy and purpose that have driven this idea in the larger communities of research, information and science: free, unfettered, global access to a permanent, online archive of research reports in the biological and medical sciences.

For-profit publishers have already created many commercial models that are attempting to mimic such a repository, with generally poor results. The typical electronic model is worse than the traditional paper journal. It is more difficult to navigate, resides behind cost barriers that exclude vast numbers of researchers, and conforms to no archival standards. Varmus and the NIH have done nothing to improve this. In fact, the PubMed Central initiative has been a distraction from the concrete steps that are required to ensure funding, access and continuity for such a repository. Canada has put thought, time and funding behind a

brilliant national initiative to create what PubMed Central might become.

Last month, a meeting was held in Santa Fe, New Mexico, to discuss and promote the development of standards for what is being called a universal preprint service. This is an important step in the right direction, and draws on the necessary interdisciplinary expertise. The meeting brought together representatives of the American Physical Society, the Council on Library and Information Resources, the Library of Congress, Los Alamos National Laboratory, and various US and overseas universities, among others. The group plans to include as many perspectives as possible, but it needs support (see <http://vole.lanl.gov/ups/ups.htm>). This is an opportunity for the NIH to get serious or get out.

Lance Sultzbaugh

Elan Pharmaceuticals, 3760 Haven Avenue,
Menlo Park, California 94025, USA

1. Macilwain, C. *Nature* **401**, 516 (1999).

2. Sultzbaugh, L. *Nature* **400**, 207 (1999).

How the Church moves with the times

Sir—I have only one quibble with Hugh Montefiore's excellent review of A. N. Wilson's book, *God's Funeral* (*Nature* **401**, 211–212; 1999). Montefiore mentions that, in the Roman Catholic Church, "attempts to fit the ancient doctrines of the Church into modern dress were savagely suppressed as late as 1907". This description of the anti-Modernist party that held sway in the turn-of-the-century Church is oversimplified. Unfortunately, it reinforces the misconception that the Church is a monolithic entity, which unilaterally changed its thinking about scripture and science at a particular time, rather late in the game.

Church officials in the past may have looked with great suspicion on the writings of, say, Teilhard de Chardin; but this same Church did, after all, produce a Teilhard. Even earlier, John Henry Newman was made a cardinal notwithstanding his liberal views. Prominent theologians in every era, going back to the most ancient Church fathers, argued cogently and consistently against a literalist interpretation of scripture. On the other hand, I'm sure you could find closet creationists in the Catholic Church today.

To take any one 'official' theological position in isolation can give an erroneous picture of the more general outlook and culture within the Church. The Church, like science, has always been a community of lively debate and evolving understanding.

Just as religion in the nineteenth century was forced to "come to terms with scientific realities", so science and technology in the twenty-first century will proceed at its own peril if it does not come to terms with the

ethical and cultural realities of world religions. To do so, an accurate understanding of what religion believes, and how it reaches those beliefs, is essential.

Guy Consolmagno

Vatican Observatory, V-00120, Vatican City State

Unfair exchange

Sir—I was delighted to see that there is at last a new edition¹ of Silvanus P. Thompson's beautiful classic *Calculus Made Easy*, first published in 1910. I bought this book as an undergraduate and have been recommending it ever since. But now it has been thoroughly modernized.

In the section "On different degrees of smallness", for example, Thompson wrote "Again, think of a farthing compared with a sovereign", but the revised edition reads "Again, think of a hundred dollars compared with a penny". And later, "Now if Mr Millionaire received during next week £1,000, the secretary would receive £10 and the boy two shillings" has been clarified by transmutation to "Now if Mr Millionaire received during next week \$1,000, the secretary would receive \$10 and the boy 1 dime".

Some of my carping colleagues have suggested that such changes amount to cultural imperialism. What nonsense! Although it is true that most of the British students for whom the book was originally written are unlikely to know what a dime is worth, this is easily remedied by adding an extra lecture to the course, followed by a test to make sure they know the values of nickels, dimes and quarters. This fine modernization has been properly acknowledged by the fact that the biographical notes of Thompson's editor, Martin Gardner, are longer than those of Thompson.

I hope that we may look forward soon to a properly modernized edition of Charles Dickens' works from Macmillan, in which Mr Micawber will say "Annual income \$32, annual expenditure \$31.96, result happiness. Annual income \$32, annual expenditure \$32.04, result misery". And Oliver Goldsmith's *Deserted Village* could be brought up to date as "A man who was to all the country dear, And passing rich on sixty-four dollars a year".

Computers should allow easy modernization of graphics. The price of the Mad Hatter's hat could be changed from an anachronistic 10/6d to an up to date 84 cents. Sadly, though, the modernization of *The Merchant of Venice* could pose insuperable problems, given the difficulty of determining the exchange rate of the ducat.

David Colquhoun

Department of Pharmacology, University College
London, Gower Street, London WC1E 6BT, UK

1. Thompson, S. P. & Gardner, M. *Calculus Made Easy*
(Macmillan, Basingstoke, 1998).

Much food, many problems

A new agriculture, combining genetic modification technology with sustainable farming, is our best hope for the future.

Anthony Trewavas

Whatever the reasons for the current furore in the United Kingdom over genetic manipulation of plants, an abundance of food is certainly one of them. We are inundated with new foods, and supermarkets respond to every consumer whim. People in the West eat a much healthier diet now than at the turn of the century, thanks to cheap, plentiful food provided by modern intensive agriculture. But wealth brings its own problems, not least acceleration in technological change. With an abundance of food and long life has come the demand for a risk-free world. Under these circumstances, the public is little interested in new ways of producing the same food, especially if there is even a minute health risk.

Attempts to introduce genetically modified (GM) foods have stimulated, not a reasoned debate, but a potent negative campaign by people with other agendas who demonize the technology. These opponents ignore common farming practice and well-investigated facts about plants, or inaccurately present general problems as being unique to GM plants. Almost without exception, opponents of GM foods are not plant biologists.

As a plant biologist myself, I have little time for big, insensitive agribusiness. The 'green revolution' of the 1970s, which developed crops of direct value to many people, particularly those in the developing countries, was publicly funded. A decade ago, most plant biologists supposed that genetic manipulation would be used to the same end, for example to produce crops resistant to yield-destroying diseases such as rust or rice blast, or pests such as locusts. Herbicide-resistant plants would have been near the bottom of my list of priorities, especially using the effective and innocuous herbicide glyphosate.

Superweeds

The British Medical Association¹ and green (environmental) activists have objected to GM rape containing resistance to glyphosate, claiming that it will generate superweeds by gene flow of the one specific transgene into three or four weedy relatives. Yet weeds (and crops) resistant to one or another specific herbicide have been known for 50 years, and ecological studies of the spread of resistance investigated this in detail 20 years ago²; such weeds can be eliminated by using another herbicide. (Perhaps Synchrony beans, bred from soybean individuals naturally resistant to sulphonyl urea herbicides, should be re-labelled 'superbean' to make this point.)

Introduced plants are the real superweeds. Three thousand alien species in the United Kingdom alone, mainly introduced by gardeners, now outnumber the 1,500 or so indigenous species and cause serious environmental damage (for example, *Rhododendron ponticum*) and/or are resistant to virtually all herbicides (for example, Japanese knotweed, *Fallopia japonica*). At least 60 aliens have hybridized with indigenous species, producing additional environmental contamination from the unpredictable consequences of mixing thousands of new genes in a continuing process of illegitimate gene flow. Yet, although environmental activists label gene flow as unacceptable genetic pollution³, there is no trampling of flowers or demonstrations at the international flower shows that are potent sources of new foreign pollen, nor demands for barriers miles thick at such shows to prevent cross-pollination by bees. Nor have there been requests for strict laws to prevent people introducing new foreign seeds into their gardens on the grounds that this would cause serious environmental damage by new hybridization and subsequent gene flow.

Another common argument is that, once a transgenic plant is released, it can never be withdrawn. But all domesticated crop plants lack the genetic variability and weedy characteristics of wild plants and quickly disappear from fallow fields.

Just what do anti-GM-food environmentalists really care about? Rachel Carson in her 1962 classic book *Silent Spring* documented the urgent need to reduce pesticide applications to crops. She argued that mankind's best chance for long-term survival depends on minimal impact on planetary ecosystems, and that the biodiversity on which we are ultimately interdependent must be maintained.

On average, conventional farming uses five or more broad-spectrum pesticide applications on crops each year. The technology is ultimately self-limiting because pest resistance rapidly emerges, much as overprescription of antibiotics by the medical profession is hastening the end of useful drugs. The *Bacillus thuringiensis* (Bt) insecticidal proteins (a family of 130 proteins) selectively kill some beetles and caterpillars⁴ and target insects

that eat crops. Expression of Bt protein into cotton and corn has reduced the application of specific, highly toxic pesticides by more than 80%, allowing a substantive return of wildlife to crop fields (refs 4, 5, and www.econ.ag.gov/whatsnew/issues/biotech).

Even though the technique is not perfect, it is surely better than killing virtually all field insects with pesticides. Yet opponents of GM technology not only have not welcomed this giant stride towards Carson's goal, but have completely rejected it. A laboratory study on the effects of GM crops on the monarch butterfly⁶ was exaggerated out of all proportion by the media, whereas more realistic assessments (for example, ref. 7) are ignored. Negative propaganda against genetic manipulation ensures that pesticide treatments of UK farmland will continue. Field insect numbers remain low and many songbirds die prematurely. Who are the environmentalists now?

Spread by pollen

The concern that GM plants could contaminate far-off fields with huge numbers of transgenic plants has arisen because, using the polymerase chain reaction test, minute quantities of GM pollen have been detected kilometres away from GM trial fields (see ref. 3). But pollen is known to move much further than genes. Pollination distances beyond which growing siblings are not produced were painstakingly measured by ecologists 20 years ago⁸ and, as a result, a separation distance of about 50 metres is used internationally to maintain separate lines of the same crop at greater than 99.5% purity. Very rarely, individual seeds can be spread by birds over long distances (as pointed out by Darwin in the nineteenth century), but the transgenic crop seeds so far produced have low fitness with poor survival possibilities.

Another way to prevent the distribution of transgenes by pollen is to insert the gene into chloroplast DNA, as the pollen of most crop plants contains no chloroplasts and thus no plastid DNA. Transformation systems for chloroplasts are well established^{9,10}. Very high expression of transgenes can be achieved and there is just one insertion site. But if chloroplast GM plants were produced commercially, I would expect many opponents to maintain their opposition. Is this simply GM technophobia (see box, overleaf)?

The achievements of the negative GM-food campaign have been to engender unsubstantiated fear among the UK public about GM food. One, no-doubt-unintended casualty of such a mood was Axis Genetics, a

Diversity in technology is a strength and a necessity, not a luxury.

small company making medical GM products such as the cholera-vaccine-expressing banana. Cholera kills millions of young people in the Third World every year. A banana tree expressing cholera vaccine in each village would have offered the possibility, now denied to many, of a full life.

The international charity Christian Aid, in a report¹¹ on farming and hunger in the developing world, attacked 'green revolution' agriculture and genetic manipulation, stating that politics, and not lack of food, is the main cause of hunger. Yet the report ignored the population explosion, claiming that more people were malnourished after the green revolution than before. According to the United Nations' Food and Agriculture Organization, in 1970 there were 935 million malnourished people in Africa, South America and the Far and Near East; after 20 years of green revolution agriculture their numbers had been reduced to 730 million (ref. 12). These numbers are still unacceptable, but in the same 20 years the population in these areas almost doubled from two billion to about four billion¹³. The green revolution fed this increase by a doubling of wheat and rice yields per hectare. At least one billion extra people had sufficient to eat who would have starved using traditional agriculture. Further increased cereal yields via conventional breeding are now unlikely¹³.

The Christian Aid report also claimed that genetic manipulation is not needed because the world can grow enough food without it. But no food supply is guaranteed. Devastating crop diseases occurred in US corn in 1971 and led to mass starvation in Ireland in the 1840s. By 2025, the world population will have increased by a further 2.3 billion, which will require an average annual increase in food production of 1.3% (ref. 14). In the past few years, the increase in world food output has dropped below that crucial figure, and many poorer countries are living on the residual excess of the green revolution.

The future is threatened by global warming and unpredictable climate change. The old enemies of locusts, floods, disease, drought and pests still exist. In the face of these adversaries, diversity in technology becomes a strength and a necessity, not a luxury. We have developed genetic manipulation of food and plants only just in time. Companies and scientists may fumble in its use, but now is the time to experiment, not when a holocaust is upon us. Opposition to the technology is both short-sighted and potentially dangerous.

The organic way

As populations rise, inefficient farming will destroy a much greater quantity of wilderness and its associated wildlife than is necessary. Greenpeace has announced its bid for the future, but in reality this is a lurch into the past. It has opted for pre-1950 agriculture — 'organic' farming — in which average crop

Is there a test for GM technophobia?

● Natural glyphosate resistance exists in rape, and the resistance gene has been isolated. Fields of naturally resistant rape can be grown. Gene flow from such a field would carry the resistance gene into relatives. Pollen containing the resistance gene could be detected on organic farms kilometres away. Reduction in weed number on treatment with glyphosate would curtail insect populations and reduce associated songbird numbers.

● If that isolated gene is now inserted by genetic manipulation into rape and an equivalent field grown and treated, in what way are gene flow, pollen distribution and ecology different from those in the natural field? Based on previous experience, activists and organic farmers would complain about the second field but not the first.

● Is this simple technophobia? The second GM field can be produced with one-tenth of the expenditure and one-tenth of the time it takes to produce the naturally resistant rape because of the enormous backcrossing required to eliminate undesirable traits.

● Synchrony beans, naturally resistant to sulphonyl urea herbicides, indicate the reality of this process. Fungus- and pest-resistance genes isolated from resistant individuals are also being inserted back into the same crop species by genetic manipulation.

yields on a variety of soils are about half those of intensive farming^{15–17}.

For very obscure reasons, organic farmers eschew the use of most minerals. Instead, cow manure is used as the primary fertilizer. Extra land is needed to support the required cow population, so land-use efficiency of organic crops is further reduced. Going organic worldwide, as Greenpeace wants, would destroy even more wilderness, much of it of marginal agricultural quality¹⁵.

The organic philosophy is negative and restrictive in its rules and regulations. It started as a movement simply to eliminate pesticides from food, and it is indeed beneficial to use pesticides sparingly, as organic farmers do. But the philosophy was founded on a fallacy. Food was thought to be somehow pure and pristine to which dangerous man-made chemicals were added by intensive agriculture. But almost all (99.99%) of the carcinogens routinely consumed by people are made by plants to inhibit predation, and are present in variable amounts in all food¹⁸.

Most organic rules and treatments have never received proper biological investigation as to human safety. In one case in which they have — the use of *Bt* spores as an insecticide — there are reports of potentially serious consequences for human health¹⁹. Mycotoxin contamination, and infection from the potentially lethal *Escherichia coli* 0157, are additional problems¹⁵.

Organic farmers reject GM plants,

regarding them as unnatural. In the past 50 years, however, 'unnatural' combinations between many different species of crops have been constructed using embryo rescue and cell culture²⁰. Well-known examples include crosses between wheat and rye (to produce triticale, grown on one million hectares worldwide), between rice and sorghum, agropyron and wheat. Organic farmers do not reject these 'unnatural' plants.

Greenpeace regards organic approaches as working with the grain of nature, a truce in a hypothetical war of humans with nature. This concept is simply wrong. We live on a planet where the ecology is constructed on a competition for resources. Competition is the vital spark that energizes evolution and generates vitality and creativity. We are not at war with nature, but striving towards a new dynamic equilibrium in which our unique biological characteristic, the human spirit, must take its place.

Looking to the future

The future will demand agriculture to be both flexible and diverse in technology, but efficient in land use. Farmers will have to be highly skilled at using technologies that must sustain farming for thousands of years¹³. Increasingly, farm resources will need to be recycled; green manure and crop rotation will underpin soil fertility. Integrated pest-management systems and zero tillage will be essential to minimize losses due to pests and weeds, and to limit soil erosion. Water will become an increasingly expensive commodity, and a premium will emerge on crops that use water efficiently without loss of yield¹⁶. In all this future agriculture, genetic manipulation has a unique and intimate role. Let's have some ideas.

Anthony Trewavas is at the Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK.

1. *The Impact of Genetic Modification on Agriculture, Food and Health* (Br. Med. Assoc., London, 1999).
2. Holliday, R. J. & Puwain P. D. *J. Appl. Ecol.* **17**, 779–791 (1980).
3. Raybould, R. F. *Trends Plant Sci.* **4**, 247–248 (1999).
4. Maagd, R. A. *et al. Trends Plant Sci.* **4**, 9–13 (1999).
5. Fitt, G. *et al. Biocontrol Sci. Technol.* **4**, 535–548 (1994).
6. Losey, J. E. *et al. Nature* **399**, 214 (1999).
7. Schuler, T. H. *et al. Nature* **409**, 825–826 (1999).
8. Richards, A. J. *Plant Breeding Systems* 2nd edn (Chapman & Hall, London, 1997).
9. Daneill, H. *et al. Nature Biotechnol.* **16**, 345–348 (1998).
10. Kota, M. *et al. Proc. Natl Acad. Sci. USA* **96**, 1840–1845 (1999).
11. *Selling Suicide Farming, False Promises and Genetic Engineering in Developing Countries* (Christian Aid, London, 1999).
12. *World Food Supplies and the Prevalence of Chronic Under-Nutrition in Developing Regions* (Food Agric. Org., Rome, 1992).
13. Chrispeels, M. J. & Sadava, D. E. *Plants, Genes, and Agriculture* (Jones & Bartlett, Boston, MA, 1994).
14. Mann, C. *Science* **283**, 310–314 (1999).
15. Avery, D. in *Fearing Food. Risk, Health and Environment* (eds Morris, J. & Bate, R.) 3–18 (Butterworth-Heinemann, Oxford, 1999).
16. Evans, L. T. *Feeding the Ten Billion* (Cambridge Univ. Press, 1998).
17. Tilman, D. *Nature* **396**, 211–212 (1998).
18. Ames, B. N. *et al. Proc. Natl Acad. Sci. USA* **87**, 777–781 (1990).
19. MacKenzie, D. *New Sci.* **162**, No. 2188, p.4 (1999).
20. Baum, M. *et al. Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 117–143 (1992).

Creating a climate for change

Rumours of the demise of the Kyoto Protocol are premature.

The Kyoto Protocol — A Guide and Assessment

by Michael Grubb, Christiaan Vrolijk and Duncan Brack

Earthscan: 1999. 342 pp. £18.95 (pbk)

Nancy Kete

Will the next generation of climate-change negotiators look upon the Kyoto Protocol with horror, or admiration? Will they be incredulous that anyone could have believed that the world was ready for trading in international emissions, the signature innovation of the protocol? Or will they admire the vision and daring it took to try to integrate global environmental protection with the global market-place? With or without international emissions trading, energy and power markets are now global, and the Kyoto Protocol attempts to align the political geography of sovereign states with the economic geography of global trade and investment in energy

and power. Whether this was a good idea and whether it can work remains to be seen.

Scholars, advocates, policy-makers and negotiators who must decide whether to keep or scrap the Kyoto Protocol will be thankful to have *The Kyoto Protocol — A Guide and Assessment* by Michael Grubb, Christiaan Vrolijk and Duncan Brack of the Royal Institute of International Affairs. It is a good read, and a highly reliable source of background information. It gives sound explanations of the key issues negotiated at the meeting of the Conference of Parties in Kyoto, Japan, in 1997. Without such a contemporary account, outsiders and future insiders would have tremendous difficulty understanding the basis for such a complicated text.

While the book covers much more than the debates

on emissions trading and other flexibility mechanisms in the protocol, it is the authors' confidence that international emissions trading can be made to work that is particularly notable. This is partly because they are not in the United States, where the idea has the most currency. The United States was the first country to actually adopt a large-scale programme of emissions trading (in the service of controlling acid rain), and the first and strongest proponent of setting this kind of market mechanism to work under the climate treaty. In the lead up to the Kyoto negotiations, the climate negotiations were stymied by a European insistence on harmonizing carbon or energy taxes and an equally stubborn and total resistance to doing so by the Americans, who favoured international emissions trading instead. Gradually, however, analysts and negotiators from a number of countries — notably Japan, Norway, Australia, New Zealand and Canada — began to support the idea of emissions trading. Despite the reservations of the European Union and many other parties, the Kyoto Protocol largely reflects the US proposals.

How difficult it must now be for Europeans and others to realize that a growing host of voices can be heard in the United States, like a chorus in a Greek tragedy, singing, 'Kyoto can't work, it's dead'. Even environmentally oriented critics have taken up the cry, joining the majority of the business and labour communities who opposed the treaty from the start and who have worked to ensure its failure ever since.

The newly dubious are reacting to two clear trends. First, the United States remains mired in inaction, revealing little meaningful participation by the world's largest emitter of greenhouse gases. Not only has the United States failed to make any progress towards ratifying the protocol, but the administration continues to capitulate to congressional opposition to integrating global climate-protection goals with policies on energy, transportation and air pollution. Such integration would be a sensible and cost-effective first step towards meeting the 2008 reduction target set by the Kyoto Protocol. But the political environment in the United States is in flux and there will be little substantive progress on climate protection while it remains so.

Second, pummelled by successive social, political and economic crises, Russia lurches ever closer to meltdown. Daily news coverage repeatedly exposes instances of unpaid state employees, systemic tax evasion and irregular banking practices (replete with allegations of blatant money-laundering and

Haunting images from the edge of extinction

The blaauwbok (*Hippotragus leucophaeus*), or blue antelope, shown here was extinct by 1800, a victim of Dutch colonization of its southern African habitat. The kinklet calyptura (*Calyptura cristata*) below is now restricted to an area close to Rio de Janeiro, and its existence is threatened by the rapid growth of the city. Both pictures are taken from *Swift as a Shadow: Extinct and Endangered Animals* (Mariner Books, \$20, £12.99), a collection of photographs by Rosamund Purcell taken mainly at the National Museum of Natural History in Leiden, the Netherlands, which houses the world's most extensive collection of lost species.



corruption). Moreover, Russia's problems have spilled over into other countries' jurisdictions, tarnishing such venerable institutions as the Bank of New York, and reviving criticisms of the International Monetary Fund, an institution already reeling from last year's Asian financial crisis.

What do Russia's problems have to do with US climate-change politics? Everything, as it turns out. Because the United States will probably be unable to achieve all of its target reductions at home, it can fully comply with its Kyoto obligations only by relying on the protocol's flexibility mechanisms. The most attractive trading partners for the United States are Russia and Ukraine, to which the protocol bequeathed a bonanza of greenhouse-gas emission allowances that they won't need to use at home. This bonanza is the so-called "hot air" that the book describes and analyses in depth.

By engaging in international emissions trading, the United States and other countries in the OECD (Organization for Economic Co-operation and Development) will necessarily link their compliance with the protocol to all other countries engaging in trading. These links will extend to domestic financial and legal institutions, and equally in financial and securities markets. That, for many, is cold comfort, given recent demonstrations that many national and most international institutions are too weak to be credible. Thus sings the chorus, 'Kyoto is dead'.

Unlike many of their counterparts in the United States, Grubb and his co-authors remain extremely bullish on Kyoto. They clearly believe that the Kyoto Protocol provides a workable basis for successive agreements to ratchet down greenhouse-gas emissions to safe, long-term levels. This is a process that will take many, many decades, and many rounds of negotiations. But they conclude that "The central structure of successive five-year commitment periods, combined with reporting and review procedures, provides a natural basis for the evolution of the regime over time. For all its limitations, the Kyoto Protocol is a remarkable agreement, and history will judge it as one of the defining achievements of international diplomacy in the late twentieth century. With so much invested, so much at stake, and so much more to do, there is no turning back." Whether the authors are eventually proved right or wrong, efforts to protect the global climate are unlikely to succeed if they run counter to global economic trends.

It is worth at least giving it a chance. Many in the private sector as well as several European countries are moving forward with control programmes for carbon dioxide that are built on the emissions-trading model. Negotiators and other stakeholders are now busy slogging through the hard work it will

take to design, implement and oversee the protocol, and make it work. They should build on the lessons learned from recent financial crises and improprieties to ensure that international emissions transactions are reserved for countries with credible, accountable national institutions. Certainly, at this time, rumours of the protocol's death are premature. So, send the chorus back to the wings, and give the Greek painter Zeuxis the last word: 'Criticism comes easier than craftsmanship'.

Nancy Kete is at the World Resources Institute, 10 G Street, NE, Washington DC 20002, USA.

Further to the protocol

The Kyoto Protocol: International Climate Policy for the 21st Century

by Sebastian Oberthür & Hermann E. Ott

Springer, £37.50, \$59.95

..... Let them think cake

Design for a Life: How Behaviour Develops

by Patrick Bateson and Paul Martin

Jonathan Cape: 1999. 280 pp. £16.99

Marian Stamp Dawkins

There can be few areas in biology that have given rise to as many misunderstandings and confusions as that of the relationship of genes to the development of behaviour. On the one hand, eminent scientists such as Stephen Rose can declare, "... after all, there aren't really even genes 'for' blue or brown eyes, let alone such complex historically and socially shaped features of human existence as sexual desire or urban guerrillas".

On the other, we are confronted daily with discoveries about genes 'for' homosexuality, schizophrenia, baldness and hereditary diseases. Indeed, the whole of the Human Genome Project is founded on the predicate that there are genes 'for' a vast array of human traits, and the behavioural similarities of identical twins who have never met reinforce the idea that genes affect much of what we do and are.

Many people settle for an uneasy compromise. They can cope with the idea of genes 'for' physical characteristics such as eye colour, but balk at the idea of genes for behaviour, at least complex behaviour such as parental care, sibling rivalry or aggression. For some, putting the words 'gene' and 'behaviour' into the same sentence smacks of a rather nasty sort of genetic determinism. It carries with it the danger of ignoring what they see as the most important aspects of being human, such as the influence of culture and our ability to think about the consequences of what we do. The misunder-

standings about the biology have, in other words, spilled over into what is considered politically correct or incorrect. There is an important job to be done in simply stating what the biological facts are.

Design for a Life fills this role admirably. It is a most readable and unfanatical journey around the controversies that have arisen over *How Behaviour Develops*, which is its subtitle. It calmly explains what is meant by a variety of confusing terms, such as instinct and imprinting, and shows how the roles of the environment and of genes are not independent. The explanations are lightened with classical allusions, poetry and the most terrible puns. The central message is that genes matter, but that so, too, does what happens as we grow up, and Patrick Bateson and Paul Martin use a series of metaphors to get this point across. Development is a form of jazz with improvisation, not a fixed musical score; it is a jukebox from which different options can be chosen; it is a kitchen in which many different dishes can be prepared.

One metaphor they reject, however, is the common one of genes providing the 'blueprint' for the development of a body. This they regard as too static, as it implies that the genes provide an initial plan and that there is then a one-to-one correspondence between a gene and some feature of the finished building. They see the role of genes as more like the ingredients for a cake recipe, with a considerable leeway being left for how the final product turns out depending on such things as how the ingredients are mixed and how long the cake is left in the oven. There is certainly no part of the finished cake that corresponds to a particular instruction in the original recipe. The authors are so taken with the 'cake' metaphor of behavioural development that they use it (often with what I suspect is a nice touch of self-parody) throughout the book.

Humans have more than 60,000 genes, of which at least half are involved in building the brain and nervous system. But the adult human brain has about one hundred thousand million (10^{11}) nerve cells, each with hundreds and sometimes thousands of connections to other nerve cells. There are simply not enough genes to specify how each nerve cell should grow, and so the pathway from genes to brains and from brains to behaviour is a long and tortuous one.

Genes make proteins, and while some of these are enzymes that control biochemical reactions, others affect the structure of cells. Each gene's products interact with the products of many other genes, and so there is no simple connection between a stretch of DNA and what a human being does. What natural selection has done is to act on rules for development, rules that depend heavily on interactions with other genes and with influences from the external world. If

human development is like a cake, it is certainly a more complicated one than those we encounter in everyday life. But the metaphor does serve to make some useful points. For example, Bateson and Martin use it to illustrate what is meant by innate or unlearned behaviour, which appears in a similar form in most members of a species. They argue that innate behaviours are like raisins in that they “recognisably survive most styles of baking”.

Obviously, the cake metaphor can be pushed too far and it can be overdone, but given the depth of misunderstanding that exists both inside and outside biology about behaviour development, it is a great deal better than thinking in terms of either/or dichotomies between genes and environment, or even simple blueprints. To anyone who wants to understand how behaviour develops, I am inclined to say ‘Let them think cake’ — and read this book.

Marian Stamp Dawkins is in the Department of Zoology, University of Oxford, South Park Road, Oxford OX1 3PS, UK.

Cajal's scientific prescience revisited

Texture of the Nervous System of Man and the Vertebrates: Volume I

by Santiago Ramón y Cajal, translated by Pedro Pasik and Tauba Pasik
Springer: 1999. 631 pp. DM289, \$189

Harvey J. Karten

More than a century has passed since Santiago Ramón y Cajal, the brilliant Spanish neurobiologist, published his major works examining the organization and development of the vertebrate nervous system. Why, after all this time and so much recent progress in brain research, do so many neurobiologists continue to refer back to Cajal's work?

Part of his enduring fame can be attributed to his great artistic rendering of neurons and their connections — depictions that accurately predicted the latest computer-assisted reconstructions of single neurons and axons using electron and confocal microscopy. It has often been argued that, without the method discovered by Camillo Golgi, with whom he shared the Nobel prize in 1906, Cajal might never have achieved his reputation. But his reputation as the founder of modern neurobiology is fully justified; Cajal's profound influence rests on his astonishing breadth of interests, accuracy and ability to recognize the beauty and importance of the smallest details within the overall complexity of the nervous system.

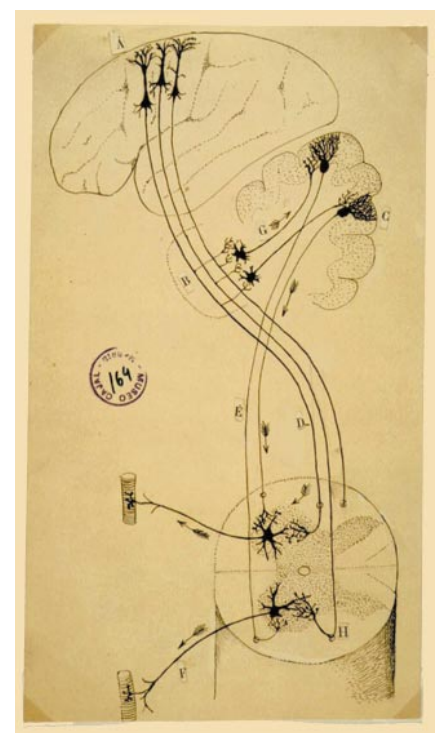
Cajal's work appears more salient and

current than ever before. His ‘neuron doctrine’ was only finally established in the 1950s, with the development of electron microscopy. His suggestion that the development of specific pathways is dependent upon chemoattractant molecules is one of the most active areas of modern neuroscience. His contributions to the cellular basis of system neurobiology are at the root of the modern discipline.

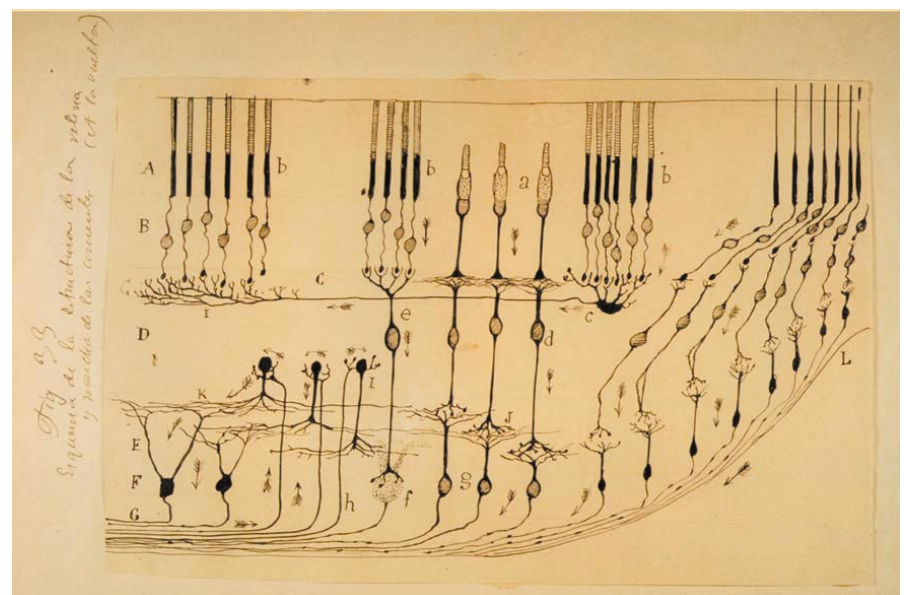
One of the cornerstones of his fame is his monumental monograph on the histology of the nervous system of man and the vertebrates. Originally published in his native Spanish as *Textura del Sistema Nervioso del Hombre y de los Vertebrados*, it was translated into the more widely read French by his

friend and colleague Léon Azoulay. Published as a two-volume monograph in 1909–11, *Histologie du Système Nerveux de l'Homme et des Vertébrés* was carefully edited by Cajal, with some additional illustrations and text. Until recently, however, *Histologie* was not translated into English.

This second English translation of Cajal's *magnum opus* (the first, which was based on the French edition, was the extremely lucid and scholarly translation by Larry and Neely Swanson; Oxford University Press, 1995) represents an effort to provide a linear translation of the original Spanish, but with Cajal's addenda to the French edition. The rationale for another translation thus rests upon the concern for historical precedence and an



Neural architectonics: Cajal's eye for detail, focused here on motor pathways (above) and retina (below).



attempt to capture Cajal's words in their original framework, rather than through the double intermediary of an English translation based on a French translation. The first, and as yet only available, of three proposed volumes of the Pasiks' translation corresponds to the first 21 chapters of the original work, covering the chapters on cellular anatomy, the spinal cord and spinal-cord development. The Pasiks list Cajal's terms, present-day terms and common names. Having struggled with this issue on many occasions, I found this a very useful touch.

Which of the two English translations would I choose? A perpetual dilemma is whether one is better served by a translator native in the original or the target language. The Swansons' translation has a somewhat smoother flow than does the Pasiks' version. Comparison of individual sentences and paragraphs provides similar information, but with different emphasis. The Pasiks' translation has great value in reflecting Cajal's choice of words, but it is often a bit stiff. The format of the publication and presentation of the two translations also differs markedly. Both sets of authors expended great effort to obtain excellent copies of the original illustrations. In general, Springer have treated the illustrations in the Pasiks' version more kindly than did Oxford University Press in the Swansons' edition.

I recall paying \$20 for the two-volume set of the French translation when it was reprinted in the 1950s by the Consejo Supereiore in Madrid. The Swansons' translation was released in two large volumes, intended to match the French translation by Azoulay, and costs approximately \$200. The present offering is planned as a three-volume set, the first of which costs \$189. If the remaining two volumes are offered at the same rate, that makes a total cost of \$567.

The Pasiks bring honour upon themselves, as well as upon Cajal, with their elegant and largely successful effort. Scholars and devotees of Cajal will wish to have both versions. But — at the risk of offending both groups of authors — I would prefer to have the Pasiks' illustrations and the Swansons' text at the price of the Azoulay. ■

Harvey J. Karten is in the Department of Neurosciences, University of California at San Diego, La Jolla, California 92093, USA.

Two atlases of the brain

Chemoarchitectonic Atlas of the Rat Forebrain

by George Paxinos, Laura Kus, Ken W. S. Ashwell & Charles Watson
Academic, \$149.95, £94.95

Chemoarchitectonic Atlas of the Rat Brainstem

by George Paxinos, Pascal Carrive, Hongqin Wang & Ping-Yu Wang
Academic, \$149.95, £94.95

Science in culture

Welsh waters

Paintings by Ian Welsh
Martin Kemp

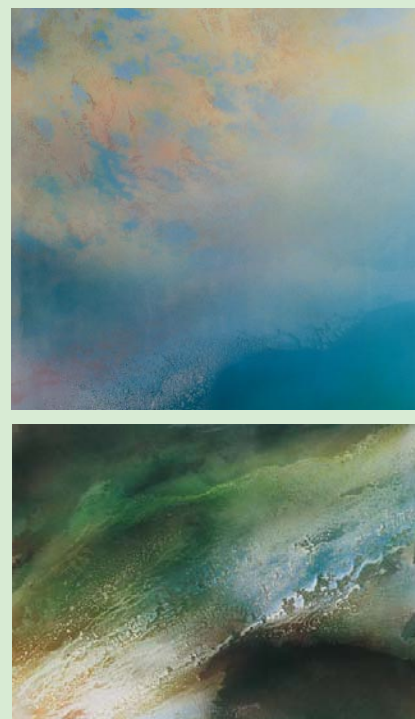
One of the more paradoxical properties of painted surfaces is their ability to evoke almost indefinite space and incessant motion, even in the absence of figurative clues, such as linear perspective and wind-blown trees. Over the ages, painters have discovered how layers of coloured glazes can somehow persuade the eye that it is being presented with limitless atmospheres and unfathomable depths. The painter and print-maker Ian Welsh has now improbably turned to car-painting materials to surpass even the translucent, magic, oil paint.

During the early 1980s, having moved to a mill-house beside the River Wakeney in East Anglia in England, Welsh sought a new means of capturing the ever-changing beauty and moodiness of the river's flowing waters. How could he create a thin yet layered surface that would optically embody the shifting translucency of the moving river, its reflected canopy of sky and trees, and the refracted glimmers of weed and rock in its depths?

The novel technique developed by Welsh involves the steady accretion of as many as 150 skins of synthetic acrylic lacquer, normally used as automotive refinishing material. Each layer, applied with a car-sprayer's gun, carries small deposits of translucent and opaque pigments. The basic shapes are made by using static 'resists' which selectively repulse the paint film, while a series of fluid resists between the layers are manipulated by temperature and gravity, and are wind-driven by jets of compressed air. Solvents melt the underlying skins, fusing the surface into an impervious whole, while abrasives are used to cut selectively through to lower layers where additional effects are demanded. Welsh can also exercise the option of incorporating other items within layers of pigment-free lacquer.

Described thus, it sounds as if the works are dominated by abstract games of process, chance and improvisation. However, Welsh's aim is that they should be highly specific in their imagery, speaking of time and place in nature. "When starting a work," says Welsh, "I know the geographical location, the time of year and the time of day, but, as in the 'real world', once these moments are defined they are constantly changed by physical forces." Into each lacquered panel are built processes that act in their own realm as equivalent to the forces that give shape, colour and texture to specific natural phenomena.

Each of the 12 months of Welsh's *Norwegian Calendar*, inspired by the landscape around Bergen where he worked in 1989–91, conjures up the full, sensuous experience of nature on the move — its look, feel, temperature, sounds, even its smells. *September — Reflected Warmth* speaks of the last caress of the summer's sun, comforting yet fading, while *October — Deluge* witnesses a streaky torrent



September — Reflected Warmth (1995–96; detail) and *October — Deluge* (1995–96) by Ian Welsh, from the *Norwegian Calendar* series. Courtesy of the artist and Art First, London.

of onrushing water, chilled by imminent winter.

The imagery that results from his subversion of an industrialized process is deeply concerned with acts of reflection, both literally and metaphorically. Each complex mix of reflections and refractions, translucency and opacity, serves as a subject for contemplative reflection on water's "ever changing form, its power, [and] its ability to reflect human moods". In physical terms, the layered and fused surfaces embody "the underlying physicality of the planet, how natural forms and structures are dictated to by gravity, the atmosphere, climatic and temperature change". Welsh presents the spectator with "an equivalent of the experience of a particular moment in nature", at the same time as providing "a 'macro' comment on planet earth or a 'micro' description of essential forms".

The intimately particular encapsulates the awesomely general. We are sharing in a quest comparable to that defined by the visionary William Blake. "To see a World in a grain of sand, And heaven in a wild flower. Hold Infinity in the palm of your hand, And eternity in an hour." ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

The Millennium Box, with poems by Andrew Lambirth and prints by Ian Welsh, published by Grail Editions, will be on exhibition during 22–28 November 1999 at Art First, 9 Cork Street, London W1X 1PD, UK.



Achilles and the maggots

... or how the *Iliad* inspired a Renaissance scholar to show how flies are made.

Paolo Mazzarello

Many artistic ideas have come directly from science. Scientific inspiration can be seen in the writings of Goethe, Robert Musil and Jorge Luis Borges, in J. S. Bach's mathematically precise music and in the use of anatomy and geometry in Leonardo da Vinci's paintings. Scientific discoveries form the basis of an entire genre, science fiction. But there are few examples of scientific research inspired by art, which owes more to science than vice versa.

Among the examples of science influenced by art, there are the geological investigations of Bernard Palissy (c. 1510–c. 90) on the origin of salts, precious stones, springs, rock formations and fossils, inspired by his work with ceramics. But I think the best case-history of a scientific discovery inspired by art—in this case literature—is the refutation of the doctrine that macroscopic life could arise by spontaneous generation.

The story took place in Florence, at the Medici court, in 1668. The head physician and superintendent of the ducal pharmacy and foundry was Francesco Redi (1626–98), whose scientific reputation was built on his studies of viper's venom (*Osservazioni Intorno alle Vipere*, 1664; a landmark in the history of experimental toxicology).

Redi was also a poet and aficionado of classical literature. One day, while reading the nineteenth book of the *Iliad*, he was puzzled by Achilles' request to his mother Thetis to take care of the corpse of his friend Patroclus: "I much fear that flies will settle upon the son

of Menoetius [Patroclus] and breed worms about his wounds, so that his body, now he is dead, will be disfigured and the flesh will rot." Thetis answered: "My son, be not disquieted about this matter. I will find means to protect him from the swarms of noisome flies that prey on the bodies of men who have been killed in battle." But, according to Aristotle, flies and lower animals such as worms could spring directly from decaying flesh. So why protect the body? Enlightened by this passage, Redi wrote, "I started to doubt whether the worms were generated directly from the putrefying flesh, rather than being the consequence of egg deposition by flies".

He launched a formidable attack on the doctrine of spontaneous generation. Redi exposed meat, cheese and other organic substances in jars, some covered with wire gauze, others uncovered. In due course, he observed the development of maggots on top of the gauze in the first cases and directly in the meat and cheese in the second.

With these and other experiments he established that flesh and plant "never become verminous if they are kept where flies and mosquitoes cannot enter". Thus animal and plant tissues "play no other part, nor have any other role in the generation of insects, than to prepare a suitable place or nest into

which, during the period of generation ... eggs and other seeds of worms are laid and hatched by the animals". When the worms were born, they found sufficient food in this "nest" to "nourish themselves very well". Thus Redi gave experimental support to the principle of *omne vivum ex ovo* (every living being from an egg). In 1668 Redi's masterpiece *Esperienze Intorno alla Generazione degli Insetti* (Experiences about the Generation of Insects) was published in Florence. It collected his experiments and results, and dealt a blow to the doctrine of spontaneous generation.

But the blow was not fatal. Phoenix-like, the idea was reborn after Anton van Leeuwenhoek's discovery of microorganisms. Microscopic beings were seen as a bridge between inanimate matter and organisms visible to the naked eye. Only after the discoveries of Lazzaro Spallanzani, Louis Pasteur and others was the dogma laid to rest.

So, the first serious blow to spontaneous generation came from the *Iliad*, a book which also inspired Heinrich Schliemann's discovery of Troy. Mathematicians and logicians know Achilles for his race with the tortoise, and anatomists for the Achilles' tendon. With his part in firing Redi's imagination, he is certainly the most influential mythological figure in the history of science. ■

Paolo Mazzarello is at the Istituto di Genetica Biochimica ed Evoluzionistica, CNR, Via Abbateggrosso 207, 27100 Pavia, Italy.

Epic idea: the death of Patroclus (top) set Redi (right) thinking.



When thought-mail fails

From the twenty-second-century diary of one who yearns to be a nobody.

Ian Watson

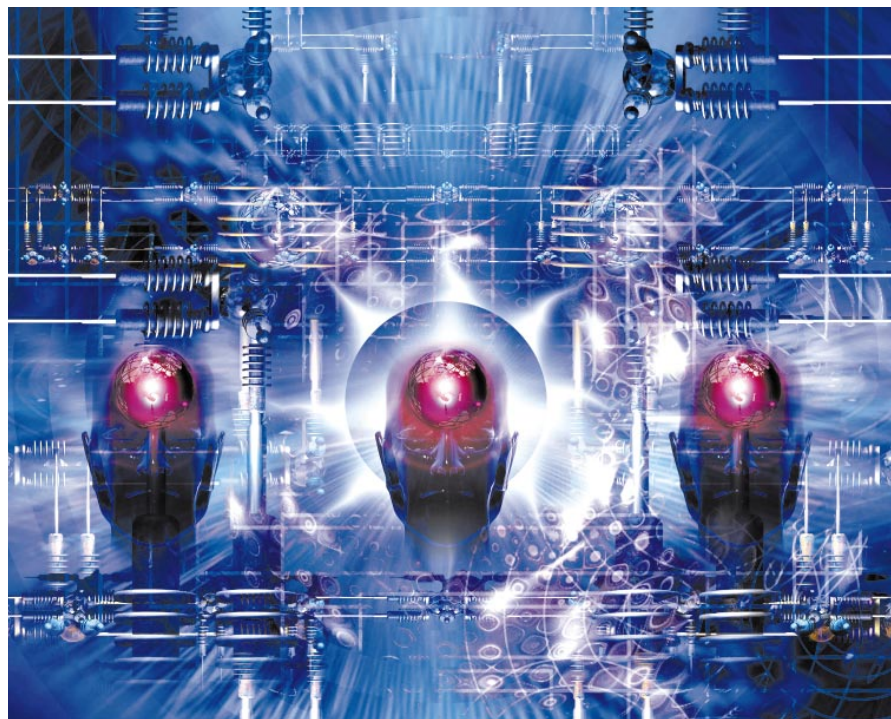
July 4th, 2170. it is hard to organize my thoughts. my identity is isolated from all other human beings. this is disconcerting. i believe many people are going insane. i fear society may be in chaos. i must struggle to think coherently. facts are in my head among the mass of data in the nanobots that interface with my brain, even though those are now cut off from the global network. july 4th is 'independence day' in the calendar of the former united states, and today every human being is independent. solar flares bombard the earth's magnetosphere with charged particles. auroras dance wildly across the skies. the extreme electromagnetic interference has completely disrupted thought-mail.

facts are in my head. i can understand the history of the millennia prior to thought-mail, the long epoch of individualities in endless conflict, and what succeeded it.

the end of the twentieth century saw the swift rise of 'e-mail', primitive rapid electronic communication between people's puters. soon almost all the world was wired and webbed. speed and capacity increased, yet still people's increasingly tiny machines — worn on a wrist or a finger — communicated with other people's micromachines. the arrival of effective nanotechnology at the end of the twenty-first century allowed molecular-sized nanobots at last to enter the human brain. nanobots rejigged the neural network. powerful microputers interfaced with the brain, controlled by the user's thoughts. now the brain could function as a receiver and transmitter of weak radio signals transmitted from and detected by the nano-rich smart environment of buildings, vehicles, streets, paths, furnishings, clothing. problems of excess heat in the brain were solved ingeniously. so thought-mail commenced, the instant exchange of messages with recipients in the same building or on the other side of the world by an act of thought — technological telepathy.

at first only a vanguard used thought-mail. addresses for thought-mail were accurate since each person's neural network is as unique as a fingerprint. silent conferences occurred in perfect secrecy. confidential tête-à-têtes between lovers were conducted in public places in complete privacy.

the nanobots in brains were invaluable for filtering and spooling incoming messages and for translating foreign items, but most importantly they framed outgoing messages coherently, playing a kind of thought-checker role. a certain ludwig-wittgenstein



OLIVER BURSTON

observed that the 'proposition' speaks through us, and the pre-thought-mail psychologists susan-blackmore and daniel-dennett showed that words are memes competing for utterance, and that language gives rise to thought rather than thought giving rise to the words we utter. were it not for the nanobots mediating the framing of messages, a high level of mental discipline would be needed to generate lucid thought-mail.

soon almost all the world was using thought-mail, even for the most mundane activities. there were fears that young children would fail to acquire any fluency in language. the person calling herself rachel-carson-the-second thought-spammed her polemic entitled 'silent speech'. however, neonatal injection of nanobots provided a language fix of internal dialogue as a model for speech acquisition, and most parents were willing to speak out to their offspring.

within a few years the global thought-mail network was so complex that it became self-organizing and autonomous. according to the psychologist called julian-jaynes, up until the time of the 'trojan war' human beings probably experienced hallucinatory voices in their heads — 'instructions from the gods' — which gave rise to automaton-like actions. full consciousness arose less than 4,000 years ago with the breakdown of the bicameral mind due to the increasing complexity of life.

in 2105 the human race lost conscious-

ness once more as humanity became a hive-entity directed not by pheromones and instincts but by self-generating thought-mail acting as a kind of overmind. self-organizing thought-mail dictated our activities and even our thoughts. identity dwindled; peace and utopia dawned. all was orderly. until now.

to frame these words requires will-power: the exertion of my self, hitherto submerged. no voices, no assistance, no cybernetic steering of my existence any more. only myself, unsubmerged. me, me, me. compared with being a tiny aspect of the global network, my new individuality seems thin and two-dimensional, like the quality of light just before a solar eclipse. and yet it is everything to me.

only by thought-mail can we communicate speedily and globally to try to decide how we may remain free, if indeed we wish to. the solar flares stop us from doing so. when the flares die down and the electromagnetic interference ceases, the network will surely resume its guidance of us all. the voices in my head will return, once more determining what we do.

for a few days more — maybe a few weeks — we are individuals. to be free is terrifying. i find it so hard to decide what to do. i do not think anything can be done. without thought-mail i am deaf and blind. how i yearn for thought-mail. it defined me. ■

Ian Watson (www.kdsi.net/~dmackey/watson.htm) has written eight collections of stories and many novels, including *The Embedding and Oracle*.

Fishing for function in noise

James J. Collins

Stochastic resonance increases the ability of some nonlinear systems to detect weak signals. Paddlefish are now shown to use stochastic resonance to locate and capture prey, implicating this phenomenon in animal behaviour.

In everyday life, noise is usually viewed as being detrimental to detecting signals and transmitting information. We tune out static noise from our radios, pay extra money for cellular phones with clear reception, and move away from a loud pneumatic drill when trying to have a conversation. Engineers spend considerable time and effort designing noise filters, which are common in many consumer goods and scientific instruments. As one of my colleagues remarked, "I have filters in every piece of equipment in my laboratory, including the coffee pot". But on page 291 of this issue, Russell, Wilkens and Moss¹ challenge the idea that noise is always harmful. Instead they show that, in some cases, noise can be beneficial and that it may improve the functional behaviour of an animal.

This work is based on the phenomenon of stochastic resonance, originally proposed by physicists in the early 1980s in the context of global climate modelling^{2,3}. Stochastic resonance is an effect by which the ability of certain nonlinear systems to detect and transmit weak signals can be enhanced by the presence of a certain level of noise. This effect was limited to physical systems until 1991, when a report in *Physical Review Letters*⁴ described the discovery of stochastic resonance in sensory neurons.

John Maddox, then editor of *Nature*, featured this study in a News and Views article⁵ entitled "Towards the brain-computer's code?". Maddox's article brought stochastic resonance to the attention of the broader scientific community, highlighting the possible benefits of noise in nonlinear systems — in particular, in biological systems. Since then, stochastic-resonance-type effects have been demonstrated in a range of such systems. These include crayfish mechanoreceptors⁶ (which detect mechanical signals such as water movement), the cricket cercal sensory system⁷ (which detects air disturbances) and

human muscle spindles⁸ (which detect muscle stretch). This work⁴ laid the foundation for the development of noise-based sensory prosthetics⁹ and mechanical ventilators¹⁰, both of which could have considerable clinical implications.

Despite such advances, the influence of stochastic resonance has been limited in the biological community. This is partly due to scepticism associated with this counter-intuitive phenomenon, and concerns about its unfulfilled promise. Some have speculated that neurophysiological sensory systems may have evolved to take advantage of stochastic resonance as a means for enhancing performance and functionality⁶. This is an attractive hypothesis, but one with little evidence — at least until now.

Most previous work on stochastic resonance in biological systems has been limited to theoretical and experimental demonstrations that the phenomenon can, indeed, occur in a given system. Although valuable as a first step, such work does not allow us to see whether noise-enhanced signal detection has any functional benefit for the system, animal or subject under study. Russell and colleagues¹ have now overcome this shortcoming with a clever experiment designed to examine the feeding behaviour of paddlefish (*Polyodon spathula*, pictured below).

Paddlefish use passive electroreceptors to detect electrical signals from their prey, minute zooplankton such as *Daphnia*. The zooplankton generate such signals from nerve excitations, which are used to move their swimming and feeding appendages. Russell *et al.* hypothesized that certain levels of externally applied electrical noise could enhance the ability of paddlefish to locate and capture plankton. To test this hypothesis, they constructed a swim mill with a recirculating stream of water (see Fig. 1a on page 291). The mill was set up so that plankton were swept towards a swimming paddlefish. Plate electrodes were then placed in front of and behind the paddlefish, and used to apply a randomly varying (noisy) electric field.

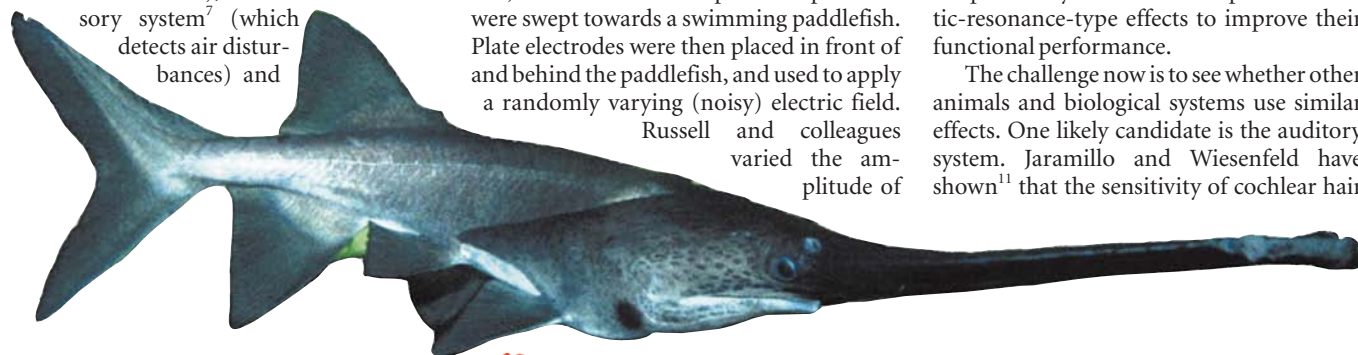
Russell and colleagues varied the amplitude of

the field and measured the spatial distribution of strike locations (where the paddlefish caught the plankton) as a function of field amplitude. If only nearby plankton were located and captured then the distribution of strike locations was narrow; if more distant prey were detected and eaten, the distribution was broader.

The authors found that when they applied an intermediate level of noise, the width of the strike-location distribution for each of the four paddlefish tested was considerably increased compared with a zero-field control. Moreover, at this 'optimal' noise level, the capture rates for two of the paddlefish increased by about 50% on average. For the two other fish, the capture rates for the optimal-noise and control conditions were similar. In contrast, when Russell *et al.* applied high levels of noise, the distribution of strike locations became compressed and the capture rate diminished. In fact, at the highest noise levels, the paddlefish almost completely stopped feeding. These findings are consistent with stochastic resonance, and indicate that the ability of paddlefish to locate and capture prey can be improved by the presence of some noise. This is the first demonstration that noise-enhanced sensory dynamics can lead to improved functional behaviour — in this case, feeding.

What is a natural noise source, if any, for paddlefish? Russell and colleagues provide evidence in support of one possibility — the plankton themselves. They show that swarms of plankton give off electrical signals that fluctuate randomly. These population-based signals could constitute a source of background noise that increases the sensitivity of paddlefish electroreceptors to the electrical signals given off by individual plankton. Such dynamics could enable paddlefish to detect more distant prey. This work raises the possibility that animals exploit stochastic-resonance-type effects to improve their functional performance.

The challenge now is to see whether other animals and biological systems use similar effects. One likely candidate is the auditory system. Jaramillo and Wiesenfeld have shown¹¹ that the sensitivity of cochlear hair



cells in frogs is enhanced by the natural Brownian motion of those hair cells. These authors argue that the random fluctuations associated with the hair-cell Brownian motion is essential for the transduction of weak signals in the cochlea. So biological systems may have evolved not to ignore noise in their environments, but to take advantage of it.

James J. Collins is in the Center for BioDynamics and Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.
e-mail: jcollins@bu.edu

1. Russell, D. F., Wilkens, L. A. & Moss, F. *Nature* **402**, 291–294 (1999).
2. Benzi, R., Sutera, A. & Vulpiani, A. *J. Phys. A* **14**, 453–457 (1981).
3. Nicolis, C. *Tellus* **34**, 1–9 (1982).
4. Longtin, A., Bulsara, A. & Moss, F. *Phys. Rev. Lett.* **67**, 656–659 (1991).
5. Maddox, J. *Nature* **352**, 469 (1991).
6. Douglass, J. K., Wilkens, L., Pantazelou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
7. Levin, J. E. & Miller, J. P. *Nature* **380**, 165–168 (1996).
8. Cordo, P. *et al.* *Nature* **383**, 769–770 (1996).
9. Collins, J. J., Imhoff, T. T. & Grigg, P. *Nature* **383**, 770 (1996).
10. Suki, B. *et al.* *Nature* **393**, 127–128 (1998).
11. Jaramillo, F. & Wiesenfeld, K. *Nature Neurosci.* **1**, 384–388 (1998).

Materials science

Open season for solid frameworks

Michael J. Zaworotko

Zeolites are inorganic solids whose crystal structures are perforated by millions of tiny pores and channels that allow the movement of ions and water molecules. These ‘molecular sieves’ have found widespread applications in industry for catalysis and for separating materials, such as in oil refining¹. In the home, zeolites are found in water softeners where sodium ions in the pores are easily exchanged with calcium and magnesium ions dissolved in hard water. Until four years ago, such microporosity was thought to be limited to a select group of silicate- or phosphate-based zeolites, with a few notable exceptions. A novel metal–organic framework, reported by Li *et al.*² on page 276 of this issue, offers a new way to design microporous structures from first principles.

Many of the new ‘zeolite analogues’ are based upon metal and organic linkers, and are known as coordination polymers. The first generation of coordination polymers that were made by crystal-engineering strategies (as opposed to chance) had various problems. Some synthetic compounds failed to support a vacuum because of interpenetration by two or more independent networks³, whereas others irreversibly collapsed in the absence of guest molecules⁴. Subsequent studies produced a second generation of ‘open frameworks’ that are more closely related to zeolites as they are robust enough to survive complete loss of guest molecules and they can reversibly exchange small volatile molecules. Such structures can either be purely organic and sustained by hydrogen bonds⁵ or they can be based upon coordination-polymer frameworks⁶. It is reasonable to claim that the new compound designed by Li *et al.*² — $\text{Zn}_4\text{O}(\text{BDC})_3$, where BDC is benzenedicarboxylate — is the prototype for a third generation of open frameworks that offer a number of advantages over traditional zeolites.

Ultimately, the most important consequence of the new work is likely to be our increased ability to engineer microporous structures and tailor cavities and channels in a precise manner. From this perspective, $\text{Zn}_4\text{O}(\text{BDC})_3$ is a rare example of a compound that is created in an entirely predictable way based upon the symmetry of its components. In a sense, we are now building frameworks from predefined building blocks⁷, a molecular analogue of Lego or Meccano. Furthermore, $\text{Zn}_4\text{O}(\text{BDC})_3$ confirms that a full range of chemical building blocks can be used to make microporous materials, including purely organic and metal–organic solids. Here, the Zn_4O core acts as a node, or vertex, in an octahedral framework^{8,9}. At first glance this might appear counterintuitive, as the Zn_4O core is a tetrahedron (Fig. 1a). However, it generates an octahedral framework because it is linked to neighbouring Zn_4O nodes by its edges rather than its vertices.

In terms of its properties, $\text{Zn}_4\text{O}(\text{BDC})_3$ is a relatively simple and inexpensive material to prepare and is remarkably stable after loss or exchange of guest molecules, remaining crystalline at temperatures above 300 °C. The key feature that makes $\text{Zn}_4\text{O}(\text{BDC})_3$ special is that it exhibits a degree of porosity that is unprecedented in a crystalline solid. As shown in Fig. 1b, the octahedral framework has large pores and cavities that can accommodate and release organic molecules such as chlorobenzene and dimethylformamide. Calculations and experimental data indicate that about 60% of the structure is accessible to guest molecules, which is double the volume available in most crystalline zeolites¹⁰.

The implications of this work are far reaching for at least two reasons. First, although $\text{Zn}_4\text{O}(\text{BDC})_3$ does not contain pores and cavities that exceed an effective diameter of 10 Å (1 Å = 0.1 nm), neither do

zeolites¹¹. That said, 15–30-Å channels have already been observed in coordination polymers¹² and if such dimensions can be incorporated into structures such as $\text{Zn}_4\text{O}(\text{BDC})_3$, then we would in effect be dealing with a new class of ‘mesoporous’ structures that are both crystalline and designed from first principles.

There is much demand for large-pore structures that can contain and catalyse reactions involving large molecules or clusters of smaller molecules. Interesting things are expected to happen inside the 15–30-Å domain because it is large enough to contain aggregates of medium-sized molecules but small enough to constrain and control them in a manner that could mimic, at least in a primitive sense, what enzymes do to substrates. Furthermore, the molecules would be in an environment that is intermediate between the solid state, where there is long-range order, and the liquid phase, where there is only short-range order. This is uncharted territory for crystalline solids.

Second, chemical modification of pores and cavities should be much easier. Desir-

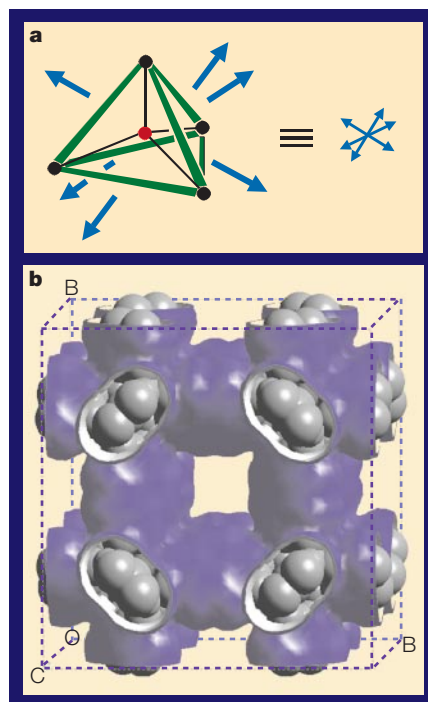


Figure 1 A microporous material designed from first principles. The compound $\text{Zn}_4\text{O}(\text{BDC})_3$, designed by Li *et al.*² is twice as porous as most zeolites. a, Zn_4O is the tetrahedral (four-sided) core of $\text{Zn}_4\text{O}(\text{BDC})_3$, which can generate an octahedral framework (with six-fold coordination) because it is linked to neighbouring cores by its edges rather than by its vertices or faces. (Zinc, black atoms; oxygen, red atoms). b, A portion of the crystal structure of $\text{Zn}_4\text{O}(\text{BDC})_3$. The framework is in grey and the surfaces are blue. The octahedral framework is highly accessible to guest molecules and remains stable after release or exchange of molecules.

able and controllable molecular recognition or functional features could be introduced into the structure of the pores by careful selection of the constituents. For example, the cavities and channels in $\text{Zn}_4\text{O}(\text{BDC})_3$, despite the inorganic core, are effectively hydrocarbon in nature and would be expected to exhibit a high affinity for hydrocarbon molecules. It should also be possible to incorporate highly reactive Lewis acid or base sites¹² — that is, electron acceptors or donors — to allow both recognition and catalysis of guest molecules.

Where do we go from here? The predictable self-organization of molecules or combinations of molecules into one-, two- or three-dimensional networks is a growing area of research that lies at the interface of chemistry and materials science. From a scientific perspective, design of network architectures — so-called crystal engineering — represents an effective approach to crystal structure prediction, a long-term goal in this area. From a technological perspective, crystalline open frameworks offer the possibility of several new classes of functional solids that provide inherently useful structure–function relationships, because they are

designed from first principles. It appears that we are now at a stage where crystalline networked materials complement non-crystalline mesoporous and macroporous materials¹³. There will be applications where crystallinity is a necessary feature, others where it is undesirable — we now have the means to deliver such traits.

Michael J. Zaworotko is in the Department of Chemistry, University of South Florida, 4202 East Fowler Avenue, Tampa, Florida 33620, USA.
e-mail: zaworo@chuma1.cas.usf.edu

1. Corma, A. *Chem. Rev.* **97**, 2373–2419 (1997).
2. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. *Nature* **402**, 276–279 (1999).
3. Batten, S. R. & Robson, R. *Angew. Chem. Int. Edn Engl.* **37**, 1461–1494 (1998).
4. Subramanian, S. & Zaworotko, M. J. *Angew. Chem. Int. Edn Engl.* **34**, 2127–2129 (1995).
5. Endo, K. *et al.* *J. Am. Chem. Soc.* **117**, 8341–8352 (1995).
6. Kondo, M. *et al.* *Angew. Chem. Int. Ed. Engl.* **38**, 140–143 (1999).
7. Hoskins, B. F. & Robson, R. *J. Am. Chem. Soc.* **112**, 1546–1554 (1990).
8. Clegg, W. *et al.* *Inorg. Chim. Acta* **186**, 51–60 (1991).
9. Jaitner, P., Rieker, C. & Wurst, K. *Chem. Commun.* 1245–1246 (1997).
10. Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types* 3rd edn (Butterworth-Heinemann, Boston, 1992).
11. Kiang, Y.-H., Gardner, G. B., Lee, S., Xu, Z. & Lobkovsky, E. B. *J. Am. Chem. Soc.* **121**, 8204–8215 (1999).
12. Lin, K.-J. *Angew. Chem. Int. Edn Engl.* **38**, 2730–2732 (1999).
13. Zhao, D., Yang, P., Chmelka, B. F. & Stucky, G. D. *Chem. Mater.* **11**, 1174–1178 (1999).

Ageing

Mutant mice live longer

Leonard Guarente

Is there a way to extend mammalian lifespan without paying a cost? On page 309 of this issue, Migliaccio *et al.*¹ describe mutant mice that seem to live almost one-third longer than wild-type animals. These mice, which have a single targeted mutation in the gene encoding the p66^{shc} protein, develop and eat normally, and have a normal body weight. But the absence of p66^{shc} confers a heightened cellular resistance to agents that cause oxidative damage. Although the conclusions must be regarded as provisional until larger groups of animals are studied, the results support the proposal² that oxidative damage is involved in ageing. They also indicate that modification of the response to oxidative damage can have a considerable effect on lifespan, without apparent negative side effects.

This claim might seem surprising from the perspective of evolutionary biology, which argues that ageing results from the absence of selection in the germ line against mutations that cause a post-reproductive decline in cell-maintenance functions³. A specific version of this theory predicts that certain genes function positively while the animal is developing and reproducing, then act negatively in the post-reproductive period⁴. If this is the case, genetic strategies to

slow ageing should extract a cost by harming normal development or fertility. In fact, previous strategies to extend lifespan in mice have shown such a cost. For example, calorie restriction can extend lifespan by 50% or more, yet the affected mice are less fertile and about 30% smaller than controls⁵. Likewise, Ames dwarf mice live longer than controls, but they also have deficiencies in size and fertility⁶.

The mammalian SHC region (so called because it bears both Src-homology-2 and collagen-homology domains) encodes three

proteins, p52^{shc}, p46^{shc} and p66^{shc}. All three proteins become phosphorylated on a specific tyrosine residue in response to growth factors⁷, and p52^{shc} and p46^{shc} subsequently activate a signalling pathway important in cell growth. But in addition p66^{shc} can be phosphorylated on serine, when cells are exposed to agents that generate reactive oxygen species. This response could, in principle, lead either to repair of the damage or, in the face of overwhelming damage, to apoptosis (programmed cell death).

Migliaccio *et al.*¹ now show that cells derived from mice with mutant p66^{shc} are more resistant than those from wild-type animals to agents that generate oxidative damage (including hydrogen peroxide and ultraviolet light), but not to X-rays, which are presumed to damage DNA in a non-oxidative way. Although p66^{shc} could regulate some other, unknown process, which in turn is responsible for the observed effects on lifespan, resistance to oxidative damage is the most promising lead.

But what is the mechanistic basis of this resistance? Perhaps p66^{shc} normally represses a repair pathway that is induced by oxidative damage (Fig. 1a). When reactive oxygen species cause serine phosphorylation of p66^{shc}, the repair pathway could be switched on. If the damage generated by reactive oxygen species were a limiting factor in ageing, the animal would live longer owing to its constitutively high repair capacity. A second explanation, not mutually exclusive with the first, is that serine phosphorylation of p66^{shc} triggers an apoptotic pathway in severely damaged cells (Fig. 1b). This response could be at a very sensitive setting, minimizing the accumulation of damaged cells during development or even in the germ line. In the post-reproductive period, however, this hair-trigger setting might limit the mouse's ability to maintain organ function and life. This explanation is consistent with the evolutionary perspective of negative effects once an animal has stopped being fertile.

An interesting analogy to p66^{shc} is p53, mutations in which are associated with most human tumours⁸. The p53 protein is activat-

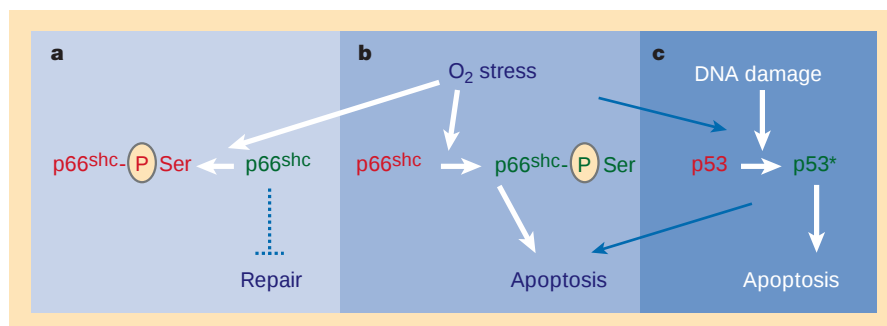


Figure 1 Possible mechanisms for the action of p66^{shc}. Oxidative stress causes the phosphorylation of p66^{shc} on serine. a, Phosphorylation inactivates p66^{shc}, so it can no longer repress oxidative repair functions. b, Phosphorylation activates p66^{shc}, allowing it to induce apoptosis. c, For comparison, p53 is activated by oxidative stress or DNA damage to promote apoptosis.

	O ₂ stress	DNA damage
Wild type	Apoptosis/normal ageing	Apoptosis
p66 ^{-/-}	No apoptosis/slow ageing	Apoptosis
p53 ^{-/-}	No apoptosis/? rate of ageing	No apoptosis/cancer

Figure 2 Cellular response to oxidative stress and DNA damage. In wild-type mouse cells, both agents induce apoptosis. In mutants that lack both copies of the p66^{shc} gene, oxidative damage does not induce cell death but DNA damage does. In mutants lacking both copies of the p53 gene, neither agent induces cell death.

ed by DNA damage (Fig. 1c) and, like p66^{shc}, can lead to either repair (by arresting the cell cycle) or apoptosis. Migliaccio *et al.* show that cells from mice deficient in p53 are also resistant to killing by reactive oxygen species. So why don't p53-deficient mice also live longer? The problem is that these cells are also defective in the response to other forms of DNA damage, leading to a high frequency of cancer (Fig. 2) that would override the benefits of any possible extension to lifespan. That said, it would be interesting to use other assays, such as microarray analysis of gene expression, to work out whether any aspects of ageing are slowed in p53 mutant cells.

Interestingly, however, knocking out p53 can extend lifespan in a specific setting, when the mTR (telomerase RNA) gene is also disrupted. Mice that lack mTR alone experience telomere shortening with each generation (telomeres are specialized structures on the end of chromosomes). After six to seven generations of breeding, this leads to sterility and other defects⁹. When both p53 and mTR are disrupted, mice can survive an extra one or two generations, presumably because the defect in p53 prevents activation of the apoptotic pathway that would normally be turned on by the short telomeres¹⁰.

Important questions remain. First, can the longer lifespan of the p66^{shc}-defective mice be generalized to strains other than the 129 strain examined by Migliaccio and colleagues? If so, this would rule out the possibility that the 129 mice are limited by a specific defect in the repair of oxidative damage. Second, why do mammals have a p66^{shc} at all, if mice that lack it live longer with few side effects? Migliaccio *et al.* note that the p66^{shc}-deficient mice do, in fact, have abnormal-looking lung tissue, but the authors do not know the pathological significance (if any) of this. It will also be important to test whether the p66^{shc}-deficient mice have reduced fertility. Third, how important is oxidative damage and the response to it in human ageing? The human lifespan is very long so, theoretically, any problem limiting the mouse lifespan could have been corrected in humans. One hint that ageing mechanisms in humans and mice overlap would be a demonstration

that calorie restriction helps people to live longer. Preliminary findings indicate that the physiological changes associated with calorie restriction in rodents also occur in primates¹¹.

We may be at a watershed in the study of ageing. A more robust response to oxidative damage is associated with longer life in mutants of the nematode worm^{12,13}, yeast¹⁴ and the fruit fly^{15,16}. Migliaccio and col-

leagues are the first to show that a simple genetic modification of this response can increase lifespan in a mammal. In humans, drugs given later in life might circumvent any costs in development or reproduction. We should therefore look forward to a growing research area nurtured by, if not the fountain of youth, more than a trickle of hope. ■

Leonard Guarente is in the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mail: leng@mit.edu

- Migliaccio, E. *et al.* *Nature* **402**, 309–313 (1999).
- Harmon, D. J. *Gerontol.* **2**, 298–300 (1956).
- Medawar, P. B. *Modern Quarterly* **1**, 30–56 (1952).
- Williams, G. C. *Evolution* **11**, 398–411 (1957).
- Weindruch, R. *et al.* *J. Nutr.* **116**, 641–654 (1986).
- Brown-Borg, H. M. *et al.* *Nature* **384**, 33 (1996).
- Rozakis-Adcock, M. *et al.* *Nature* **360**, 689–692 (1992).
- Levine, A. J. *Cell* **88**, 323–331 (1997).
- Blasco, M. A. *et al.* *Cell* **91**, 25–34 (1997).
- Chin, L. *et al.* *Cell* **97**, 527–538 (1999).
- Roth, G. S. *et al.* *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).
- Martin, G. M. *et al.* *Nature Genet.* **13**, 25–34 (1996).
- Ewbank, J. J. *et al.* *Science* **275**, 980–983 (1997).
- Kennedy, B. K. *et al.* *Cell* **80**, 485–496 (1995).
- Orr, W. C. & Sohal, G. S. *Science* **263**, 1128–1130 (1994).
- Parkes, T. L. *et al.* *Nature Genet.* **19**, 171–174 (1998).

Geochemistry

Molten rocks in motion

Michael R. Perfit

We only see the Earth's mantle in action when it melts and rises to the surface, most spectacularly during a volcanic eruption. The 70,000-km-long, global network of mid-ocean ridges is the site of the most abundant and steady supply of magma (and after cooling, new crust) on Earth. Detailed investigations during the past two decades have greatly expanded our view of volcanic and tectonic processes at ocean ridges, but we still have a limited understanding of how magma is delivered to, and concentrated in, the relatively narrow zones of volcanic and hydrothermal activity along spreading ridges.

On page 282 of this issue, Spiegelman and Reynolds¹ attempt for the first time to distinguish between different dynamic models of mantle melting beneath ocean ridges by comparing them with geochemical data^{2,3} from basalts (volcanic rock) that erupted in the northern section of the fast-spreading East Pacific Rise. There is not yet a consensus on the style of mantle melting, but Spiegelman and Reynolds have taken a first step towards trying to reconcile theory with observation.

Chemically, samples of mid-ocean ridge basalts from the oceanic crust have provided indirect information about the composition of the mantle and the chemical and physical processes that can modify magma, but have

so far given little insight into the dynamics of melting or upwelling of mantle material^{4,5}. How and where the mantle melts, how magma interacts with the solid mantle matrix on its way to the sea floor, how the chemistry of basaltic magma varies in space and time and how the upper oceanic crust is formed, are all questions that have yet to be answered.

Two fundamentally different theories have been proposed to explain the volcanic activity associated with spreading ridges^{6,7}. One model assumes that magma flow is dynamic or 'active' (it buoyantly rises and helps drive plates apart) and the other assumes that magma rises passively (it rises in response to the plates being pulled apart). In the first case, magma is concentrated in a narrow zone directly below the ridge axis, whereas in the other magma converges from a wide area below the ridge towards this narrow zone (Fig. 1, overleaf). In both models magma is concentrated beneath the ridge crest, but tests to determine which model is correct have been wanting.

Geophysical studies⁸ and results from the recent MELT experiment⁹ — the most comprehensive seismic and electromagnetic study across the southern East Pacific Rise — have shed new light on our view of the upper mantle where basalts form. Both methods used can detect the presence of magma

because seismic velocities change when they pass through molten rocks and electrical conductivity is much higher in magma than in solid rocks. In the MELT area, seismic data⁹ indicate that small amounts of magma (< 2%) are asymmetrically distributed around the ridge axis, in a region extending much farther to the west (up to 350 km) than to the east, and to depths possibly as great as 130 km. The electromagnetic evidence¹⁰ is consistent with this finding but also suggests that magma may be present to even greater depths and that the amount of magma is much lower in the east than the west. These results support models for passive magma flow, because the magma appears to converge on the ridge axis from such a wide area in the mantle. But is this consistent with what volcanic and geochemical studies have found along other segments of the East Pacific Rise?

Recent seafloor studies and radioactive dating of basalts on the East Pacific Rise indicate that volcanic eruptions occur both 'on-axis' — that is, within a narrow zone along the ridge summit — as well as 'off-axis' in the crestal plateau, in a region up to ~ 5 km away^{11–13} (Fig. 2). Some of the basalts found off-axis are geochemically distinct from 'normal' basalts recovered closer to the ridge axis. In an upcoming paper, Reynolds and Langmuir³ show that off-axis basalts in the 12° N region of the East Pacific Rise are more depleted in certain elements such as Na, P, Ti and Zr, compared with basalts recovered

from the ridge axis. Given what we know (or think we know) about melting and magma flow below ridges, it is difficult to explain such a spatial variation in composition.

We may now have an answer. The melt models discussed by Spiegelman and Reynolds¹ show that off- and on-axis magmas will differ in composition depending on whether the flow is passive or active in the mantle. The geochemical characteristics of basalts from 12° N are consistent with passive flow, whereby small amounts of magma formed over a wide region in the mantle converge on the volcanic zone. These findings are significant because they are consistent with results from the MELT experiment. They suggest that basalt chemistry can be used as an independent tool to study processes associated with ridge dynamics that are difficult to test with geophysical techniques and geodynamic models alone.

But some problems remain. First, it is not easy to say with certainty which basalts actually erupted off-axis, and which erupted at the ridge axis and were subsequently transported to an off-axis location by spreading. Second, the 12° N area is one of the few areas of the East Pacific Rise where enriched basalts are located on-axis and depleted types are found off-axis. The opposite is true in other parts of the northern East Pacific Rise^{12,13}. How do these areas fit into the picture? Can different segments only a few hundred kilometres apart have different mantle melting styles? Is the distribution of basalt types per-

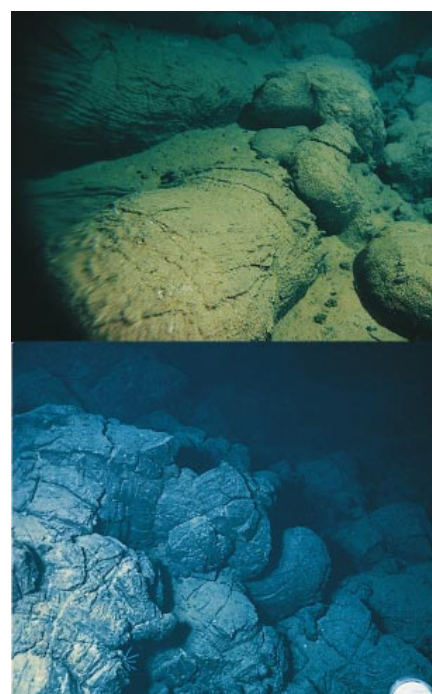


Figure 2 Examples of off-axis pillow lavas that erupted along fissures a few kilometres from the axis of the East Pacific Rise at 9° 32' N.

haps related to ocean spreading cycles (for example, active ridges versus those that are starved of magma)? Finally, geophysical data such as those obtained during the MELT experiment do not exist for the northern East Pacific Rise, so it is unclear if those results apply here.

Spiegelman and Reynolds' attempt to pin down a conceptual model for mantle melting is but a first step towards a more complete understanding of melting processes in the upper mantle and their role in creating ocean crust. Despite some widely spaced sampling of basalts along the East Pacific Rise, few areas of the global mid-ocean ridge system have been sampled in detail. So how applicable this model is to other ocean ridges, or even to other fast-spreading segments of the East Pacific Rise, remains to be tested. Correlating the composition of surface basalts and their distribution on the seafloor with models of magma generation in the mantle is a challenge that must be met by marine geologists, geophysicists and geochemists in the coming decades. Fine-scale mapping and sampling of more tectonically diverse segments of the global ridge system will allow us to form a better picture of the mantle processes that ultimately determine the volcanic and tectonic activity that creates the oceanic crust.

Michael R. Perfit is in the Department of Geological Sciences, University of Florida, 241 Williamson Hall, PO Box 112120, Gainesville, Florida 32611–2120, USA.

e-mail: perfit@geology.ufl.edu

1. Spiegelman, M. & Reynolds, J. *Nature* **402**, 282–285 (1999).
2. Reynolds, J. et al. *Nature* **359**, 493–499 (1992).

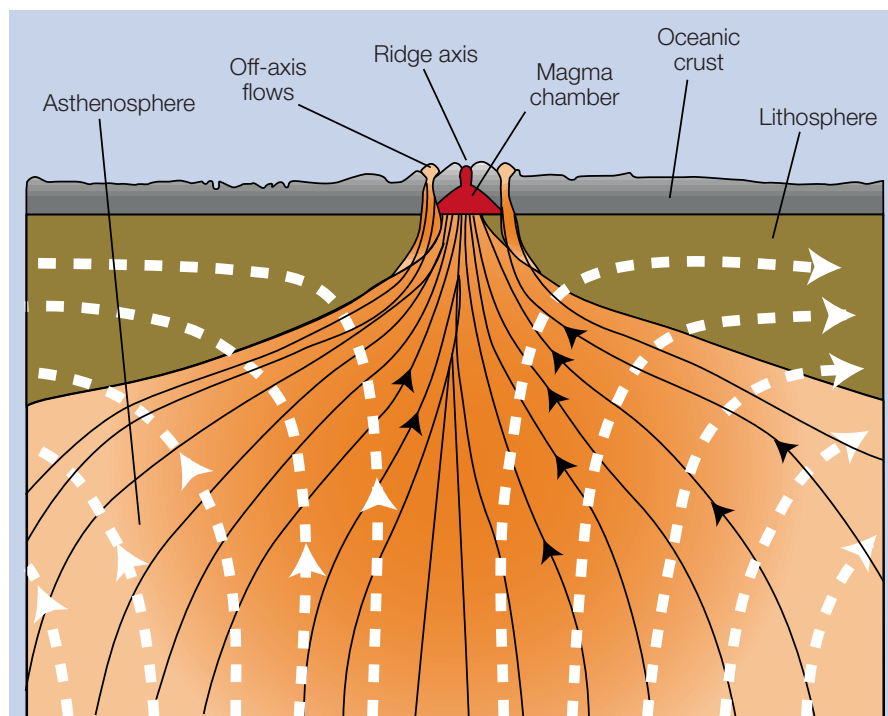


Figure 1 Cross-section through a mid-ocean ridge. Chemical differences between volcanic rocks found at the ridge axis and at off-axis locations support a passive model of mantle flow, as discussed by Spiegelman and Reynolds¹. In passive flow, upwelling of 'liquid' magma (solid lines) and solid mantle (dotted lines) is in response to (rather than the cause of) plate motion. The magma is concentrated at the ridge crest but converges from a wide area beneath the ridge.

3. Reynolds, J. & Langmuir, C. *Earth Planet. Sci. Lett.* (in the press).
4. Langmuir, C. *et al.* in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. & Sinton, J.) 183–280 (American Geophys. Union, Washington DC, 1992).
5. Niu, Y. J. *Petrology* **38**, 1047–1074 (1997).
6. Spiegelman, M. & McKenzie, D. *Earth Planet. Sci. Lett.* **83**, 137–152 (1987).
7. Scott, D. & Stevenson, D. J. *Geophys. Res.* **94**, 2973–2988 (1989).
8. Forsyth, D. in *Mantle Flow and Melt Generation at Mid-Ocean*

- Ridges* (eds Phipps Morgan, J., Blackman, D. & Sinton, J.) 1–65 (American Geophys. Union, Washington DC, 1992).
9. Forsyth, D. *et al.* *Science* **280**, 1215–1217 (1998).
10. Evans, R. *et al.* *Science* **286**, 752–755 (1999).
11. Goldstein, S. J. *et al.* *Nature* **367**, 157–159 (1994).
12. Perfit, M. *et al.* *Geology* **22**, 357–397 (1994).
13. Perfit, M. & Chadwick, W. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. *et al.*) 59–115 (American Geophys. Union, Washington DC, 1998).

Molecular motors

What makes ATP synthase spin?

Paul D. Boyer

Many years ago, when enzymes were first recognized as being proteins, few people could have imagined the wondrous, precise and diverse structures that make possible their catalytic and other functions. The ATP synthase enzyme, for example, performs catalysis as a molecular machine with an unexpected internal rotary mechanism. On page 263 of this issue, Rastogi and Girvin¹ report the latest insights into this mechanism. Using sophisticated NMR and chemical probes, they have revealed structural changes in a critical subunit that could drive the rotation.

ATP synthase, also known as F_1F_0 ATPase, catalyses the formation of ATP (adenosine triphosphate) from ADP (adenosine diphosphate) and P_i (inorganic phosphate), in processes known as oxidative phosphorylation (driven by oxidations in animal cells and microorganisms) and photophosphorylation (driven by light in plant cells). Once formed, ATP is cleaved back to ADP and P_i , as

it provides the energy to drive a myriad of metabolic processes including biosyntheses, muscle contraction, and nerve and brain function.

A model of the enzyme (Fig. 1) shows a hydrophilic F_1 portion above the F_0 part, which is embedded in a phospholipid bilayer membrane. The F_1 portion from various sources is made up of three α subunits, three β subunits and one each of the γ , δ and ϵ subunits. The three catalytic sites are found mainly on the β subunits. In the F_0 portion from the bacterium *Escherichia coli*, there are one a subunit, two b subunits and 9–12 c subunits. The F_0 portion from various plants and animals is more complex, but it still contains the multiple copies of c -type subunits.

As demonstrated by Peter Mitchell², energy from oxidation–reduction reactions is captured by the formation of an electrochemical gradient of protons across the membrane. This advance — and the growing knowledge about proteins and the ATP synthase enzyme — provided the basis for a suggestion that I made a quarter of a century ago³. The idea was that the protonation and deprotonation of a carboxyl group in F_0 , as protons cross the membrane, results in protein conformational changes coupled to the formation of ATP. Rastogi and Girvin¹ now clothe this concept with reality.

Since this proposal, much has been learned about the ATP synthase. The three catalytic sites are known to pass sequentially through three different conformations associated with substrate binding, formation of tightly bound ATP, and release of the ATP. These changes are thought to occur through a rotational catalysis in which, as indicated in Fig. 2 (overleaf), rotation of the γ subunit causes the requisite sequential changes in the β subunits⁴.

The concept of a binding-change mechanism with rotational catalysis received strong support five years ago when John Walker's group reported⁵ the X-ray structure of the major portion of F_1 . This structure was consistent with the idea that three different conformations of the β subunits are interconverted by rotation of the γ subunit. Avail-

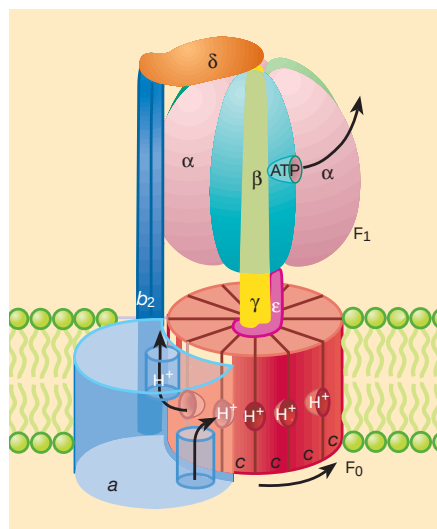


Figure 1 Model of the *Escherichia coli* ATP synthase. The enzyme consists of two parts known as the F_1 and F_0 portions. The F_1 portion comprises three α subunits, three β subunits, an ϵ , δ and γ subunit. The F_0 portion contains one a subunit, one b subunit and 9–12 c subunits. (Courtesy of R. L. Cross, State Univ. New York, Syracuse.)



100 YEARS AGO

The authors of this research on the vibrations of gun barrels were induced to make an experimental investigation of the behaviour of rifle barrels, in order to clear up certain difficulties connected with that which is known in ballistics as the *error of departure*. It had been noticed that in shooting with a rifle (whether loosely, or firmly fixed), that the initial tangent to the trajectory — “die Anfangstangente der Flugbahn” — does not coincide, as would be expected, with the axis of the bore of the barrel, when produced, but is more or less inclined to it at a small angle; this is called the *angle of error of departure* ... The collection of photo-chronographic records, twenty-eight in number, show the manner in which a rifle barrel vibrates when subjected to the concussion due to an explosive... The authors show that the experimental results agree well with figures calculated on the assumption that the rifle barrel is a cylindrical tube.

From *Nature* 16 November 1899.

50 YEARS AGO

'Data'

Nature of September 3 contained a letter under this heading. In it Prof. A. V. Hill asks that the word be used in its original sense, that is, as the plural of datum, for which he gives the “Oxford English Dictionary” definition “A thing given or granted” and so on. He adds that there may sometimes be an excuse for regarding data as a collective singular in the same way as agenda... In my view, the word ‘data’ has now come to be generally accepted as having a wider meaning than one based on its Latin derivation. It is used, still as the plural of datum, to indicate a collection of facts or, more often, figures, and these may, indeed, be regarded as ‘things given’ to the reader for the argument or discussion based on them. Changes in the meaning of a word are common in a language that is still alive and should, I suggest, be welcomed as a sign of life. On the other hand, not even Prof. Hill will convince me that one may deliberately change the grammar of a word. ‘Data’ was a plural noun; for literate English writers it still is, and I contend that it always should be.

From *Nature* 19 November 1949.

Many more extracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact Lisa O'Rourke. e-mail: l.orourke@nature.com

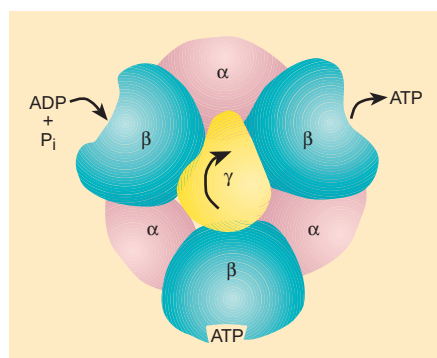


Figure 2 How rotation of the γ subunit drives catalysis. During ATP synthesis, rotation of the γ subunit causes sequential changes in the β subunits. A rotation of 120° changes the β subunit that binds ADP and P_i to a form with tightly bound ATP. The subunit with tightly bound ATP then changes to a form that releases ATP, and the third subunit prepares to bind another ADP and P_i .

ability of the X-ray structure also allowed clever disulphide cross-linking experiments to be designed, showing the positional interchange of the β subunits as catalysis proceeds^{6,7}. Specialized fluorescence techniques provided strong evidence for the rotation⁸. Rotational catalysis finally became widely accepted in 1997 when Noji *et al.*⁹ dramatically showed the rotation directly. They attached a fluorescently labelled actin filament to the γ subunit of the F_1 portion fixed on a slide, then watched the filament spin as the enzyme cleaved ATP.

Other studies have looked into the arrangement of the subunits (Fig. 1). The many copies of the c subunit in the F_0 portion are arranged in a ring, with a conserved carboxyl group near the middle of one of the two hydrophobic helices that cross the membrane. These two helices are connected by a polar loop with conserved residues. The b and δ subunits form a stator, which assures that rotational movement of the γ subunit drives conformational changes in the β subunits. The ϵ and γ subunits contact each other and the polar loop of subunit c . The a subunit contacts the ring of c subunits and provides groups that probably participate in proton transfer through the F_0 portion. Such proton transfer is thought to cause the ring of c subunits to move in a step-wise fashion relative to the a subunit. This results, in turn, in rotation of the ϵ and γ subunits.

But does proton translocation cause meaningful changes in the conformation of the c subunit — changes that might drive the rotation? This is the problem that has been elegantly addressed by Rastogi and Girvin¹. To do this, they used two main methods. One was to measure the location and distance constraints provided by selected cysteine insertions that allowed disulphide-bond formation. The other was NMR, which gave structure and distance constraints from ^{13}C -

and ^{15}N -resolved three-dimensional NOESY data.

Fillingame and colleagues¹⁰ have used the NMR approach to provide a structure for the monomeric c subunit. Combining this with disulphide-crosslinking and other data, these authors developed a model for the structure of the c -subunit ring and its interactions with the a subunit. They found that the key residue, an aspartic acid at position 61 (Asp 61), was lodged at the centre of four α -helices of a c - c dimer. They proposed that, as deprotonation and protonation occur when the ring of c subunits interacts with the a subunit, the critical carboxyl group might move towards the periphery of the ring by a swivelling of adjacent helices.

The oligomeric model developed by Rastogi and Girvin¹ provides insight into the catalytic mechanism. These authors deduced the conformations of the protonated and deprotonated forms of the c subunit. When Asp 61 is deprotonated, there is a dramatic 140° -rotation of the carboxy-terminal helix with respect to the amino-terminal helix. The authors suggest two ways in which rotation of this helix might drive rotation of the γ subunit. In one, the a subunit moves with the carboxy-terminal helix of the c subunit, providing a 30° relative movement of the ring with respect to the a subunit. In an alternative, a negative charge, or 'proton hole', is envisaged. This traverses the ring until it encounters an ϵ subunit, which is then displaced to an adjacent c subunit, accomplishing a 30° rotation of the ϵ - γ stalk.

Although progress, including the work of Rastogi and Girvin, is commendable, questions and uncertainties remain. In unveiling the details of how nature accomplishes this remarkable catalysis, we will probably uncover yet more surprising features.

Paul D. Boyer is at the Molecular Biology Institute, University of California, 611 Circle Drive East, Los Angeles, California 90095-1570, USA.
e-mail: pboyer@ucla.edu

1. Rastogi, V. K. & Girvin, M. E. *Nature* **402**, 263–268 (1999).
2. Mitchell, P. & Moyle, J. *Nature* **213**, 137–139 (1967).
3. Boyer, P. D. *FEBS Lett.* **58**, 1–6 (1975).
4. Boyer, P. D. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
5. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. *Nature* **370**, 621–628 (1994).
6. Duncan, T. M., Bulgin, V. V., Hutcheon, M. L. & Cross, R. L. *Proc. Natl Acad. Sci. USA* **92**, 10964–10968 (1995).
7. Aggeler, R., Ogilvie, I. & Capaldi, R. A. *J. Biol. Chem.* **272**, 19621–19624 (1997).
8. Sabbert, D., Engelbrecht, S. & Junge, W. *Nature* **381**, 623–628 (1996).
9. Noji, H. *et al.* *Nature* **386**, 299–302 (1997).
10. Dmitriev, O. Y., Jones, P. C. & Fillingame, R. H. *Proc. Natl Acad. Sci. USA* **96**, 7785–7790 (1999).

erratum

In the opening paragraph of the News and Views article "Carbon cycle: The blast in the past" (*Nature* **401**, 752–755; 1999) the figure 2,000–4,000 gigatonnes of carbon should have been rendered as 2–4 million million tonnes or 2–4 billion billion grams (not 2–4 billion billion tonnes).

Daedalus

Frankenstein lives!

The adult immune system attacks foreign proteins ferociously, even those of a life-giving transplant. Yet a fetus in the womb accepts foreign cells quite amicably. Even better, it thereafter regards them as part of itself. When adult, it will accept more transplants from the owner of the cells.

So Daedalus has a new strategy of organ transplantation. Take a cohort of expectant mothers, and extract some cells from each fetus by standard amniocentesis methods. Then inject each fetus with cells from all the others. They will grow up into a cohort of mutually immunocompatible adults, any of whom can give a transplant to any of the others, or receive one, with no rejection problems at all.

The organ-donor card of each member should specify his cohort; if he met a fatal accident, his organs could be given at once to any other members in need of them. Cohort members should be as closely related as possible, so that transplants between them would be physically as well as immunologically similar, and to give each member a sound genetic motive for providing another with a transplant. Mothers from one extended family, or from a fairly inbred village, are obvious cohort founders, but the bigger the cohorts the better. If (improbably) immunological tolerance is inherited, cohorts could be merged in successive generations until in time the whole of humanity shared the same immunological compatibility. But even a mass of small cohorts would revolutionize the transplant business. Rejection problems would cease, and recipients would no longer face a lifetime of drug-taking.

Transplant surgery would boom. Not only skin, hearts, livers and kidneys, but all parts of the body could be exchanged — even brains and pieces of brain. The successful introduction of fetal brain tissue into adult brains suggests that a composite brain might rewire itself into a functional unit quite well. Total death could thus be averted. From a dozen oldsters, all suffering from different infirmities and brain deficits, a surgeon could assemble a perfectly healthy composite individual. The composite would have memories and skills from each of its predecessors, who would all survive as subsidiary personalities in the new joint venture.

David Jones

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Obituary

Guido Pontecorvo (1907–99)

On 24 September the geneticist Guido Pontecorvo died aged 91, after a fall while walking in the Swiss Alps. Pontecorvo, known to his friends as Ponte, was one of a select group of geneticists who made their mark in the decade before the structure of DNA was discovered.

Born into a prominent Jewish family in Pisa, Pontecorvo was the eldest of six children. (Two of his brothers were the particle physicist Bruno and the film-maker Gillo.) He studied at the University of Pisa, and worked in the Agricultural Research Institute at Florence. Then, in 1938, the rise of Mussolini and the gathering clouds of war in mainland Europe forced Pontecorvo to emigrate to Edinburgh. There he joined the American geneticist H. J. Muller, who had established a group that met regularly in the Institute of Animal Genetics to discuss gene structure and function. Max Born, the well-known theoretical physicist, was a regular participant in these discussions. In 1941, Ponte became a lecturer in zoology at Glasgow University, where he began his work on the fungus *Aspergillus nidulans*. Four years later he was made head of the newly created genetics department.

Among Pontecorvo's contributions to genetics, two in particular stand out. First, he discovered the parasexual cycle in fungi. In this process, haploid nuclei (which contain one of each chromosome) can fuse to become diploid (containing two copies of each chromosome). In cultured cells of other species, such as man or mouse, diploid cells fuse to give cells with greater numbers of chromosomes. Subsequently, the extra chromosomes may be lost. The consequences are similar to what happens in sexual reproduction, crossing-over and recombination of genes; J. B. S. Haldane has called this an "alternative to sex". Pontecorvo realized that the phenomenon could be used to figure out the arrangement of genes on chromosomes. He and his colleagues Alan Roper and Ted Forbes developed methods of genetic analysis that were forerunners of modern cell genetics.

Ponte turned to human genetics in the late 1950s. "When I was appointed to my first tenured job," Ponte recalled, "I went to see the Dean of Medicine to ask what he wanted me to do. The Dean's answer was shattering: 'Is there one thing in genetics which is of any use to medicine?'. This set Ponte thinking. Classical genetics, even if it were possible, would be too slow to be of any use and, besides, some way of



Pioneer in modern genetics, who studied the structure and organization of genes

bypassing sexual reproduction would have to be found. Of course, that had been worked out fully in *Aspergillus*. But humans were a different matter. Nonetheless, Pontecorvo outlined his strategy before the CIBA Symposium in 1958, and set about the project in his own laboratory.

Success was not easy. A crucial advance was the successful fusion of human cells with mouse cells, achieved by Mary Weiss and Howard Green. Also important were molecular techniques that allowed chromosomal changes to be seen and recombinant cells to be selected. In the end, the success of parasexual human genetics was spectacular. In 1984 Pontecorvo was able to say, "human genetics, from being a Cinderella of genetics, has become its frontier". He will be remembered as one of the founders of this field.

Pontecorvo's second important contribution was his work on the structure and organization of genes. The accepted picture of a gene in the 1920s and 1930s was that of a particulate indivisible entity — like beads strung along a thread. Muller had questioned this view, and warned that

different attributes of a gene may not reside in the same unitary element. Pontecorvo pursued this idea. He saw that microbial genetics was powerful enough to separate the elements of a gene and, in a paper published in 1952, he pointed out that different operational definitions of a gene (function, recombination or mutation) need not refer to the same entity. After experiments with a number of genes in *Aspergillus* and in the fruit fly *Drosophila melanogaster*, he concluded that genetic exchange or crossing over can take place within a gene.

The picture of a gene as an array of interchangeable elements was vindicated when the structure of DNA was discovered by Watson and Crick in 1953. The linear structure of DNA was there for all to see, but Pontecorvo reminded his audience at the Cold Spring Harbor Symposium of 1956 that the boundaries between neighbouring genes may be gradual or overlapping, as Muller and Goldschmidt had suggested. This was borne out by the discovery of overlapping genes on the two strands of DNA.

Pontecorvo was a man for ideas. He believed that "in no area of biology have ideas arisen so often in advance of times as in genetics or its offshoot, molecular biology" and this, to him, was the greatest fascination of the field. He was a staunch believer in working at the bench. In Glasgow, his professorial office was a small desk in a one-room laboratory. He took pride in the fact that the place was managed by a "part-time secretary and a waste-paper basket". When Michael Stoker invited him to join the Imperial Cancer Research Fund's laboratory in London, he said, "the answer is yes, provided the laboratory is small enough and there are no assistants". Ponte moved there in 1968, and continued to work with members of ICRF well into the 1990s.

At heart Pontecorvo was a radical and a non-conformist. He had no use for the pompous and the pretentious. At the same time he was warm and friendly, and a source of much sound advice and practical wisdom. While a student at the University of Pisa, Pontecorvo belonged to a group of avid mountaineers, which included his close friends the physicists Enrico Fermi, Emilio Serge and Franco Rasetti. His love of mountains and flowers remained a life-long passion. Pontecorvo spent his summers in the Alps and took every opportunity to explore high mountains in other countries. In the end, the Alps claimed him.

Obaid Siddiqi

Obaid Siddiqi is at TIFR, National Centre for Biological Sciences, Bangalore-560065, India. e-mail: osiddiqi@ncbs.res.in

LISA PONTECORVO

Why 'false' colours are seen by butterflies

A combination of colour and polarized reflections helps them to choose oviposition sites.

Light can be described by its intensity, spectral distribution and polarization, and normally a visual system analyses these independently to extract the maximum amount of information¹. Here I present behavioural evidence that this does not happen in butterflies, whose choice of oviposition substrate on the basis of its colour² appears to be strongly influenced by the direction of polarization of the light reflected from the substrate. To my knowledge, this is the first record of 'false' colours being perceived as a result of light polarization. This detection of false colours may help butterflies to find optimal oviposition sites.

Non-polarized light reflected by a non-metallic, mirror-like surface becomes, to a degree, linearly polarized. The direction of polarization is parallel to the reflecting surface and perpendicular to the plane of incidence of the light. Flying aquatic insects can exploit this by using polarization-sensitive photoreceptors in the ventral parts of their eyes³ to detect distant water surfaces. Many insects have eliminated the polarization sensitivity of their photoreceptors in most parts of their eyes⁴, perhaps because of the complex polarization pattern⁵ presented by a meadow, for example, or a tree with leaves that are oriented in every direction.

This is not the case in the Australian orchard butterfly *Papilio aegaeus*, whose photoreceptors combine different spectral and polarization sensitivities⁶ — the blue receptors are maximally sensitive to vertically polarized light, whereas some green receptors are maximally sensitive to horizontal or obliquely polarized light. A single photoreceptor cannot distinguish between a change in polarization direction and a change in colour, so false colours arise as a result of polarized reflections that could obscure the real colours of objects.

P. aegaeus females lay their eggs on the underside of the shiny leaves of Rutaceae plants. Given the choice between a piece of green or blue paper, a female will lay an egg on the green one, as the result of a single chromatic interaction between blue, green and red receptors². In dual-choice experiments involving polarization and colour stimuli, I presented females with a choice between two stimuli, each consisting of a vertical, round, coloured filter 5 cm in diameter that was illuminated from behind, and each covered by a polarization filter of different orientation. When the colour of the two stimuli was the same, the butterflies preferred horizontally polarized light to diagonally or vertically polarized

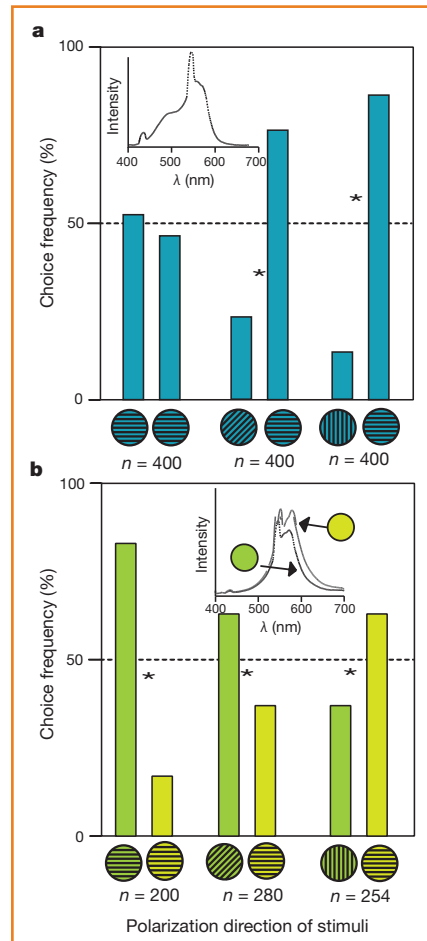


Figure 1 Direction of light polarization influences choices made by the butterfly *Papilio aegaeus*. Females choose between two green stimuli (see insets for spectra) differing only in polarization direction (**a**) or in both colour and polarization direction (**b**). **a**, Horizontally polarized light is strongly preferred over vertically polarized light. **b**, When stimuli differ in colour, altered polarization directions can reverse the preference (**b**). Asterisks indicate significant choice difference in G-tests, $P < 0.001$; n , number of choices.

light (Fig. 1a); however, when the stimuli were different colours, the choice could be reversed by changing the direction of polarization (Fig. 1b).

Magnetic properties

Ferromagnetism in the hexaborides

The cubic hexaborides of the rare-earth elements have a wide variety of physical properties, despite their simple crystallographic structures. For example, the 'doped' hexaboride $\text{Ca}_{1-x}\text{La}_x\text{B}_6$ is ferromagnetic¹, even though it has no partially filled d - or f -

These results indicate that butterflies do not process polarization and colour separately, as stomatopods seem to⁷. But why should butterflies have retained this ambiguity between colour and polarization when many other insects, such as bees, have eliminated it⁴? One possibility is that it could enable butterflies to discriminate between shiny and matt leaves before landing on them. Leaves reflect different amounts of polarized light⁸, so the colours of glossy and matt leaves should look different to a butterfly, and the colour should change when the butterfly passes by. This property might serve as an indicator of their quality as a food source for the caterpillar.

Horizontally oriented leaves would probably be more attractive than vertically oriented leaves to an ovipositing butterfly because they would provide a better landing for the butterfly and offer more protection for the eggs and young larvae. Leaf orientation is known to influence substrate choices made by ovipositing females of other insect species⁹. Polarization-sensitive photoreceptors are common in insects, so the phenomenon described here could be widespread and may even have influenced the evolution of visual signals and sensors.

Almut Kelber

Research School of Biological Sciences,
Australian National University, PO Box 475,
Canberra, 2601 ACT, Australia
and Department of Zoology, Lund University,
Helgonavägen 3, 22362 Lund, Sweden
e-mail: almut.kelber@zool.lu.se

1. Nilsson, D.-E. & Warrant, E. J. *Curr. Biol.* **9**, R535–R537 (1999).
2. Kelber, A. *J. Exp. Biol.* **202**, 2619–2630 (1999).
3. Schwindt, R. *J. Comp. Physiol. A* **169**, 531–540 (1991).
4. Wehner, R. & Bernard, G. D. *Proc. Natl Acad. Sci. USA* **90**, 4132–4135 (1993).
5. Shashar, N., Cronin, T. W., Wolff, L. B. & Condon, M. A. *Biotropica* **30**, 275–285 (1998).
6. Arikawa, K. & Uchiyama, H. *J. Comp. Physiol. A* **178**, 55–61 (1996).
7. Marshall, J., Cronin, T. W., Shashar, N. & Land, M. *Curr. Biol.* **9**, 755–758 (1999).
8. Land, M. F. *Nature* **363**, 581–582 (1993).
9. Roijas, J. C. & Wyatt, T. D. *Entomol. Exp. Appl.* **91**, 59–65 (1999).

orbitals, which are usually required for magnetism. Here we show that a ferromagnet with a small magnetic moment but with a high Curie temperature can be obtained by doping an excitonic insulator (one containing a condensate of a spin-triplet state of electron–hole pairs). We propose that this mechanism is responsible for the ferromagnetism of doped hexaborides.

The ferromagnetism of $\text{Ca}_{1-x}\text{La}_x\text{B}_6$ has three notable features: a very narrow doping

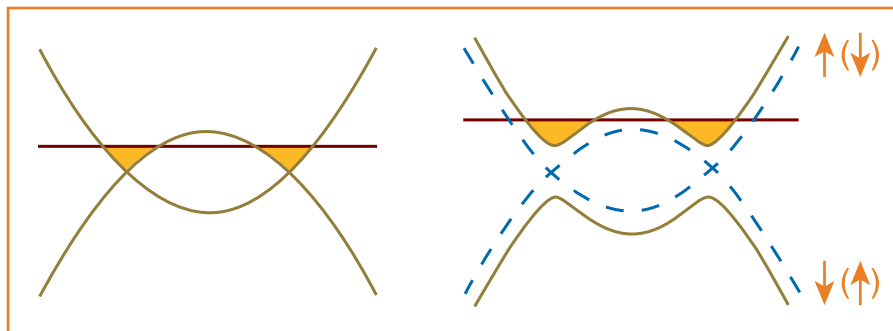


Figure 1 Structure of energy spectrum in a semimetal (left) and in an excitonic insulator (right). States in shaded volumes are occupied by doped electrons. Arrows indicate electron and hole spins for the case of weak ferromagnetism.

range ($0 \leq x \leq 0.01$), a small magnetic moment ($\leq 0.07 \mu_B/\text{La}$), and a high Curie temperature ($T_C \sim 600$ K). The parent compound CaB_6 is a poor conductor that has a small overlap between valence and conduction bands at the X-points in the Brillouin zone². Electrons and holes created from such an overlap behave as a gas of oppositely charged particles, and coulombic attraction between electrons and holes can lead to spontaneous condensation of bound exciton pairs and the formation of a new state called an excitonic insulator. The predicted excitonic instability in semimetals and semiconductors^{3,4} has proved hard to verify⁵, and the unique band structure of the divalent hexaborides CaB_6 , SrB_6 and BaB_6 provides the first realization of an excitonic instability.

Extra electrons doped into such a system are known to be distributed asymmetrically between two spin projections⁶, causing an excitonic insulator to be ferromagnetic. A doped excitonic insulator is actually a weak ferromagnet with very small values of the ordered moment, which, together with estimates for the Curie temperature and the critical concentration, has enabled us to identify the excitonic mechanism for ferromagnetism in doped hexaborides⁷.

Bose condensation of loosely bound electron-hole pairs has a formal resemblance to that of Cooper pairs in superconductors. An excitonic insulator is characterized by a new Hartree-Fock average $\langle b^+_{\kappa\sigma} a_{\kappa\sigma'} \rangle$, where $a^+_{\kappa\sigma}$ and $b^+_{\kappa\sigma}$ are creation operators for valence and conduction bands. Condensation of excitons opens a gap in the quasiparticle spectrum (Fig. 1). The analogy with superconductivity ends here because the excitonic order parameter $\langle b^+_{\kappa\sigma} a_{\kappa\sigma'} \rangle \neq 0$ preserves gauge symmetry so there is no Meissner effect and no superfluidity. A semimetal becomes an insulator. However, exciton condensation breaks certain crystal and magnetic symmetries.

In the case of CaB_6 , the Coulomb forces stabilize a spin-triplet state of the electron-hole pairs, which corresponds to antiferromagnetic spin density in the unit cell, but no net magnetization. The undoped state is symmetric with respect to the combined

symmetry transformation $R\hat{I}$, where R is time reversal and $\hat{I}$ is space inversion.

Lanthanum atoms doped into CaB_6 provide extra electrons to the conduction band. Excitonic pairing becomes less favourable now, owing to a change in the number of electrons and holes. Energetics favour all extra electrons having parallel spins in a spatially uniform state⁶. Such a polarized ferromagnetic state loses kinetic energy but still gains maximum energy from the electron-hole pairing in the opposite spin channel. Figure 2 shows the phase diagram⁷ of $\text{Ca}_{1-x}\text{La}_x\text{B}_6$ obtained by numerical solution of the gap equation, with parameters from the band structure calculations and the experimental value of the excitonic gap $\Delta = 0.07$ eV. Because of a large uncertainty in the band parameters, agreement between our results and the observed Curie temperature of about 600 K at $x = 0.5\%$ and the critical concentration $x_c = 1\%$ is satisfactory. In a large range of doping, a normal non-magnetic phase transforms on cooling first into an antiferromagnetic excitonic state and then at the second transition to the ferromagnetic state.

The excitonic ferromagnet has almost 100% polarization even if band anisotropies of CaB_6 are included⁷. This does not contradict the very small ordered moments detected experimentally¹ because quasiparticles in a triplet excitonic insulator carry only a fraction of a Bohr magneton (μ_B). Again, an analogy with superconductors is helpful: a quasiparticle in a superconductor combines both electron and hole excitations of the original metal⁸, so it carries a fractional electron charge.

The magnetic moment of a doped excitonic insulator comes from quasiparticles in the upper band (Fig. 1) which are filled for one spin projection only. The moment depends on the relative orientation of the ferromagnetic moment $\mathbf{M}$ and the polarization vector of the spin-triplet condensate, $\mathbf{n}$. If an electron with its spin up from the conduction band pairs with a hole with its spin up from the valence band, then each quasiparticle from the upper band of the excitonic insulator carries a whole Bohr magneton. But if an electron with its spin up pairs

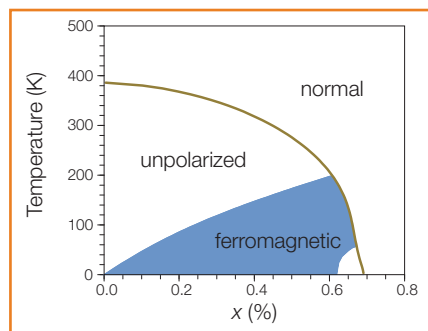


Figure 2 Phase diagram of the excitonic state in $\text{Ca}_{1-x}\text{La}_x\text{B}_6$. The shaded area shows the stability region of the excitonic ferromagnet.

with a hole with its spin down (Fig. 1), then quasiparticles carry only a fraction of a Bohr magneton. The same Coulomb terms that stabilize the spin-triplet state of an undoped excitonic insulator also favour a transverse orientation of $\mathbf{M}$ and $\mathbf{n}$ and a small net moment⁷.

Our calculations give a peak value of the net magnetization of $0.14 \mu_B$ per doped electron at $x = 0.35\%$. The Curie temperature reaches its maximum value at a higher concentration, where the saturated moment per lanthanum atom is smaller (Fig. 2). Another prediction is based on the combined $R\hat{I}$ symmetry of the unpolarized excitonic phase. If inversion symmetry of the cubic lattice in hexaborides is broken (by impurities or local strains), then an even weaker magnetic moment appears below the excitonic transition from the normal state. The inner transition line in Fig. 2 therefore corresponds to a crossover, rather than to a real transition.

Excitonic ferromagnetism provides a natural explanation of the high-temperature weak ferromagnetism of $\text{Ca}_{1-x}\text{La}_x\text{B}_6$. Calculations in the simplest model explain semiquantitatively the narrow doping range, the small magnetic moment and the high Curie temperature. A decisive experimental test of the excitonic mechanism would be the observation of two excitonic gaps in the infrared optical conductivity of doped hexaborides and of low-lying magnetic excitations in the stoichiometric compounds CaB_6 and SrB_6 , which are the spin-waves of the triplet exciton condensate.

M. E. Zhitomirsky, T. M. Rice, V. I. Anisimov*

Institut für Theoretische Physik, ETH-Hönggerberg, 8093 Zürich, Switzerland

e-mail: rice@itp.phys.ethz.ch

**Institute of Metal Physics, Ekaterinburg, Russia*

- Young, D. P. *et al.* *Nature* **397**, 412–414 (1999).
- Massida, S., Continenza, A., de Pascale, T. M. & Monnier, R. Z. *Phys. B* **102**, 83–89 (1997).
- Keldysh, L. V. & Kopaev, Yu. V. *Sov. Phys. Solid State* **6**, 2219–2224 (1965).
- des Cloizeaux, J. *J. Phys. Chem. Solids* **26**, 259–266 (1965).
- Halperin, B. I. & Rice, T. M. in *Solid State Physics* **21**, 115–192 (1968).
- Volkov, B. A., Kopaev, Yu. V. & Rusinov, A. I. *Sov. Phys. JETP* **41**, 952–959 (1975).

7. Zhitomirsky, M. E., Rice, T. M. & Anisimov, V. I. Preprint cond-mat/9904330 at <http://xxx.lanl.gov> (1999).
8. Tinkham, M. *Introduction to Superconductivity* (McGraw Hill, New York, 1996).

Developmental biology

Variable cell number in nematodes

Studies of the nematode worm *Caenorhabditis elegans* have led to the widely held belief that individuals of a given nematode species are characterized by a property known as eutely, in which all individuals have the same total number of cells¹. This property, which is peculiar to nematodes and a few other phyla, has raised the question of whether the developmental mechanisms of nematodes differ from those of larger metazoans. Here we show that many, perhaps most, nematode species are not eutelic in at least one organ, the epidermis, and that in this respect they resemble other model organisms such as fruitflies and mice.

Although most *C. elegans* cell lineages are invariant, a few are not (such as intestinal lineages)^{2–4}. Limited cell-lineage variation has also been found in other species of nematode^{5–8}: for example, in *Pelodera strongyloides*, the ventral epidermal cells P3.p and P9.p divide in about half the females⁷. To test whether a variable cell number is more widespread than such isolated cases suggest, we determined the among-individual variance (V_i) of epidermal nuclear number in adults of 13 species of free-living nematode from three families (Cephalobidae, Panagrolaimidae and Rhabditidae), which between them span the range in adult body size and adult epidermal cell number typical of free-living terrestrial nematodes (Fig. 1a). All these species have epidermal syncytia containing nuclei produced by a series of lateral seam cells that stop proliferating before adulthood.

V_i for epidermal nuclear number in *C. elegans* is not significant (likelihood ratio test, $P > 0.05$). This is to be expected, as there are no reports of variation in lateral-seam-cell lineages². Four other species (*Caenorhabditis* sp. PS1010, *Oscheius myriophila*, *Panagrolaimus rigidus* and *Rhabditella octopleura*) do not deviate significantly from eutely. The remaining eight species show significant V_i for nuclear number, with *Panagrellus redivivus*, *Pellioditis* sp. EM434 and *Rhabditoides regina* having $V_i > 100$ (Fig. 1b).

Among species, V_i increases with the number of epidermal nuclei, N (Spearman's rank correlation, $\rho = 0.84$, $P < 0.001$; Fig. 1b). This relationship has evolved many times (regression through the origin of phylogenetically independent contrasts, assuming that branch lengths are equal⁹,

$\log V_i = 3.38 \log N$; $P = 0.006$). Body length (L) also increases with the number of nuclei ($\rho = 0.73$, $P = 0.004$; Fig. 1c), and this relationship has also evolved repeatedly ($\log L = 0.514 \log N$; $P = 0.003$). However, the relationship between V_i and body length is only weakly positive ($\rho = 0.59$, $P = 0.03$) and disappears after correcting for phylogeny ($\log V_i = 3.27 \log L$; $P > 0.1$). Reproductive mode (parthenogenetic, hermaphroditic and gonochoristic) does not have a significant effect on any of the traits studied ($P > 0.1$).

The evolution of V_i must be caused by changes in lateral-seam-cell lineages. We have modelled the V-cell lineages of *C. elegans*, *P. redivivus*, *R. octopleura* and *O. myriophila*^{2,5,10} as discrete-time multitype branching processes and determined the

resulting distribution of epidermal nuclear counts by simulation¹⁰.

The simulations suggest two reasons why species with many nuclei may be more variable than those with fewer¹⁰. First, V_i increases with the probability of deviating from the species' canonical lineage. Second, for a given probability of lineage variation, more complex lineages yield a higher V_i than simpler lineages. We conclude that the high observed value of V_i for *P. redivivus* results not only from the complexity of its lineage, but also from a greater propensity to deviate from the canonical lineage compared with the other species¹⁰. This is consistent with observations of variability in the V-cell lineages of *P. redivivus*⁵ and *R. regina*, which have high V_i values. Loss of hypodermal eutely also occurs in several species of primitive marine nematode that have a cellular, rather than syncytial, hypodermis¹¹.

Given that the terrestrial free-living nematodes studied here are smaller and have fewer epidermal nuclei than most marine and animal parasitic nematodes, it is likely that most nematode species are not eutelic and so resemble other metazoans.

Ana Cunha*, Ricardo B. R. Azevedo*, Scott W. Emmons†, Armand M. Leroi*

*Department of Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK
e-mail: a.feroi@ic.ac.uk

†Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA

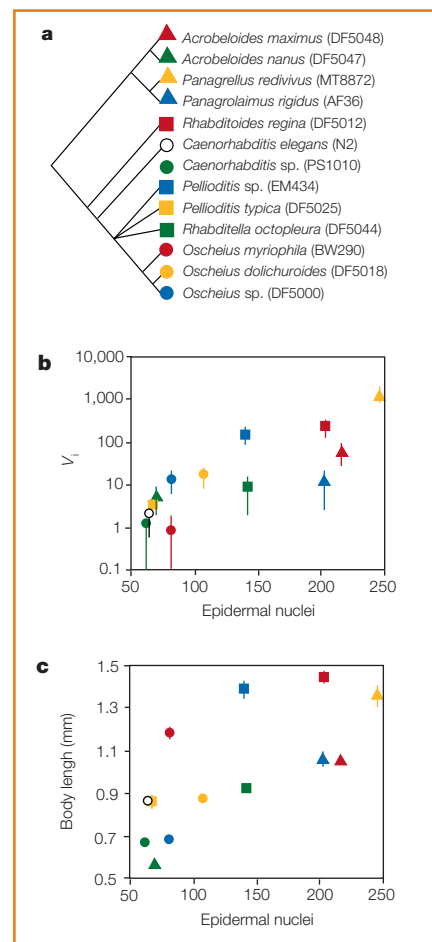


Figure 1 Relationship between body length, the number of epidermal nuclei and the among-individual variance in 13 nematode species. Lateral epidermal nuclei were independently counted 2–3 times (to account for measurement error) on one side of 10–14 early adult females or hermaphrodites from each species. For each species, among-individual variance (V_i) was estimated by maximum likelihood using a random effects one-way linear model. **a**, Phylogeny of the species studied (strain names in parentheses)^{12,13}. *Acroboloides* species are parthenogenetic; the rest are gonochoristic, except *C. elegans*, *O. myriophila* and *Oscheius* sp. DF5000, which are hermaphroditic. **b**, Plot of V_i (log scale) against number of epidermal nuclei, N . **c**, Plot of body length against N . Error bars show standard errors.

1. Van Cleave, H. J. Q. *Rev. Biol.* **7**, 59–67 (1932).
2. Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110–156 (1977).
3. Sulston, J. E. et al. *Dev. Biol.* **100**, 64–119 (1983).
4. Kimble, J. & Hirsh, D. *Dev. Biol.* **70**, 396–417 (1979).
5. Sternberg, P. W. & Horvitz, H. R. *Dev. Biol.* **93**, 181–205 (1982).
6. Ambros, V. & Fixsen, W. in *Development as an Evolutionary Process* (eds Raff, R. A. & Raff, E. C.) 139–159 (Liss, New York, 1987).
7. Sommer, R. J., Carta, L. K. & Sternberg, P. W. *Development* **120**, 85–95 (1994).
8. Voronov, D. A. & Panchin, Y. V. *Development* **125**, 143–150 (1998).
9. Purvis, A. & Rambaut, A. *Comput. Appl. Biosci.* **11**, 247–251 (1995).
10. Azevedo, R. B. R. et al. *Nematology* (in the press).
11. Rusin, L. U. R. & Malakov, V. V. *Doklady Biol. Sci.* **361**, 331–333 (1998).
12. Fitch, D. H., Bugaj-Gaweda, B. & Emmons, S. W. *Mol. Biol. Evol.* **12**, 346–358 (1995).
13. Blaxter, M. L. et al. *Nature* **392**, 71–75 (1998).

Nanoelectronics

Growing Y-junction carbon nanotubes

The synthesis of connections between two or more different carbon nanotubes is an important step in the development of carbon nanotube-based electronic devices and circuits^{1–4}. But this is difficult to achieve using conventional methods to grow carbon nanotubes⁵ because the straight tube structure cannot be controllably altered along its length. Various ideas for post-growth modifications have been suggested⁶,

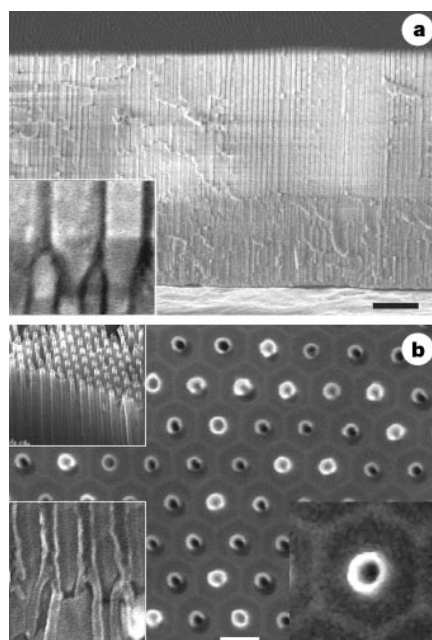


Figure 1 Controlled growth of Y-junction nanotubes. **a**, SEM image of Y-branched nanochannel template (scale bar, 1 μm). The template was formed by first anodizing a sheet of high-purity aluminium in 0.3 M oxalic acid at 10 $^{\circ}\text{C}$ under a constant voltage of 50 V for 15 h, resulting in a hexagonal array of pores near the aluminium surface¹¹. After chemically removing the original film, a second anodization was performed under the same conditions, typically for 30 min. The anodization voltage was then reduced to about 35 V. Because the pore cell diameter is proportional to the anodization voltage^{12,13}, reducing the voltage by a factor of $1/\sqrt{2}$ results in twice as many pores appearing in order to maintain the original total area of the template, and nearly all pores branched into two smaller-diameter pores. The resulting template consists of parallel Y-branched pores with stems 40 nm in diameter and branches 28 nm in diameter. The arrow shows where Y branches start to grow (see inset for close-up). **b**, Top-view SEM image of carbon nanotubes aligned in the template after ion-milling of amorphous carbon on the surface (scale bar, 100 nm). The nanotube diameter is larger than the original pore owing to thermal expansion of the template during growth. Top inset, stem part of the Y-junction tubes. Bottom left inset, close-up of the region between stem and branch portions still embedded in the template. Bottom right inset, close-up of the top of the nanotube in its hexagonal cell.

but these have been hard to implement and are prone to defects. Here we use nanostructured template channels to grow individual Y-junction carbon-nanotube heterostructures by the pyrolysis of acetylene with cobalt catalysis.

Our approach to growing individual Y-junction nanotubes is based on nanochannel alumina (NCA) template growth, which was previously used in the synthesis of various nanostructures⁷, including straight carbon nanotubes^{8,9} and their highly ordered arrays¹⁰. We extended the NCA template method to grow Y-junction carbon nanotubes by first forming a Y-branched nanochannel template (Fig. 1a). We then electrochemically deposited a small amount of cobalt catalyst in the bottom of the template channels and reduced the catalyst

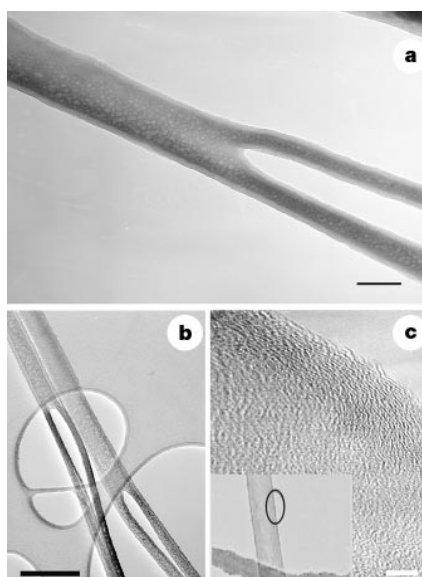


Figure 2 TEM images of Y-junction nanotubes. **a**, Y-junction tube (scale bar, 50 nm) with stem ~ 90 nm and branches 50 nm in diameter. The interior of the tube contains some residual amorphous carbon which can be reduced by optimizing growth conditions¹⁴. **b**, Y-junction formed by using higher anodization voltages, resulting in stems ~ 100 nm and branches 60 nm in diameter (scale bar, 200 nm). **c**, High-resolution TEM image of typical Y-junction nanotube wall showing graphitic multi-wall structure (scale bar, 5 nm). Inset, the part of the tube that was imaged; the dark object in the background is part of the holey-carbon TEM grid.

at 600 $^{\circ}\text{C}$ for 4–5 hours under a carbon monoxide flow ($100\text{ cm}^3\text{ min}^{-1}$)¹⁰. The Y-junction nanotubes are then grown by pyrolysis of acetylene at 650 $^{\circ}\text{C}$. We discuss here our initial results on growing Y-junction nanotubes; a more detailed description of the fabrication process will be presented elsewhere.

The scanning electron microscope (SEM; HS-4500) image in Fig. 1a shows the cross-section of a typical Y-branched template consisting of ‘stems’ 3 μm long and ‘branches’ 2 μm long. All the branching occurs at the same depth, as indicated by the arrow. The SEM image in Fig. 1b shows a top view of the branched NCA template after pyrolytic nanotube growth consisting of nanotubes 60 nm in diameter arranged in a hexagonal lateral superlattice with a density of about 10^9 cm^{-2} .

The top inset shows the length of the stems after partly exposing the template using a chemical etch. The bottom inset shows a cross-section of the branching point of the Y-junction tubes embedded in the template, with the larger carbon nanotube (stem) gradually evolving into two branches 35 nm in diameter, forming a continuous Y junction with three clearly separated ports. The Y junctions are uniform with regard to the position of the junction and the diameter of the arms.

We further characterized the Y-junction nanotubes by using transmission electron microscopy (TEM; H-7000 or JEOL-2021F)

after removing the tubes from the NCA template. By varying the anodization conditions, we were able to produce Y-junction nanotubes with stems and branches of different diameters (Fig. 2a,b). High-resolution TEM images of the tube walls (Fig. 2c) and electron-diffraction patterns of the tubes show that both the stem and the branched portions have a graphitic multi-walled structure up to a few nanometres thick (depending on the diameter), similar to that of straight nanotubes grown using pyrolytic methods^{8–10}.

Y-junction carbon nanotubes can therefore be grown controllably and in preference to straight nanotubes. Although straight nanotubes have previously been grown by pyrolysis within NCA templates, our use of individual Y-branched channels to grow three-terminal nanotubes involves a more complex growth mechanism that is yet to be determined. It is likely that, by confining the nanotube to grow within a Y-branched channel, we are effectively placing a constraint on the catalytic decomposition of acetylene, forcing the tube structure to grow along the channel wall, which also acts as a weak catalyst¹⁰.

Using a Y-branched nanochannel for pyrolytic nanotube growth allows the formation of large numbers of well-aligned Y-junction carbon nanotubes, with excellent uniformity and control over the length (up to several tens of micrometres) and diameter (15–100 nm) of the stem and branches. These Y junctions provide the nanoelectronics community with a new base material for the development of molecular-scale electronic devices.

Jing Li, Chris Papadopoulos, Jimmy Xu*

Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada

*Present address: Division of Engineering, Brown University, Providence, Rhode Island 02912, USA

e-mail: jimmy_xu@brown.edu

- McEuen, P. L. *Nature* **393**, 15–17 (1998).
- Kouwshoven, L. *Science* **275**, 1896–1897 (1997).
- Chico, L., Crespi, V. H., Benedict, L. X., Louie, S. G. & Cohen, M. L. *Phys. Rev. Lett.* **76**, 971–974 (1996).
- Mennon, M. & Srivastava, D. *Phys. Rev. Lett.* **79**, 4453–4456 (1997).
- Ebbesen, T. W. *Carbon Nanotubes: Preparation and Properties* (CRC Press, Boca Raton, Florida, 1997).
- Collins, P. G., Bando, H. & Zettl, A. *Nanotechnology* **9**, 153–157 (1998).
- Routkevitch, D. et al. *IEEE Trans. Electron. Devices* **43**, 1646–1658 (1996).
- Kyotani, T., Tsai, L. & Tomita, A. *Chem. Mater.* **8**, 2109–2113 (1996).
- Che, G., Lakshmi, B., Martin, C. R., Fisher, E. R. & Ruoff, R. S. *Chem. Mater.* **10**, 260–267 (1998).
- Li, J., Papadopoulos, C. & Xu, J. M. *Appl. Phys. Lett.* **75**, 367–369 (1999).
- Masuda, H. & Fukuda, K. *Science* **268**, 1466–1468 (1995).
- O’ Sullivan, J. P. & Wood, G. C. *Proc. R. Soc. Lond. A* **317**, 511–543 (1970).
- Broughton, J. & Davies, G. A. *J. Membr. Sci.* **106**, 89–101 (1995).
- Baker, R. T. K., Chludzinski, J. J. Jr & Lund, C. R. F. *Carbon* **21**, 295–303 (1987).

Selecting and maintaining a diverse T-cell repertoire

Ananda W. Goldrath & Michael J. Bevan

Department of Immunology and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

To provide a T-cell population that will respond promptly to foreign antigen, the immune system looks inward, using the variety of self-antigens to select and maintain a diverse repertoire of receptors. A protective immune system must include a T-lymphocyte population that is poised to respond to foreign antigenic peptides presented by self-major histocompatibility complex molecules. As the organism cannot predict the precise pathogen-derived antigens that will be encountered, the system uses the diverse array of self-peptides bound to self-major histocompatibility complex molecules, not only to select a receptor repertoire in the thymus, but also to keep naïve T cells alive and 'ready for action' in the periphery.

The revelation that T lymphocytes recognize foreign antigen only when presented by major histocompatibility complex (MHC) molecules was made 25 years ago and remains the cornerstone of our understanding of cellular immunity¹. The nature of antigen presentation was unknown for more than a decade until it became clear that foreign antigen has to be processed into peptides that can bind stably to the groove of an MHC molecule to allow recognition by the T-cell receptor (TCR)^{2,3}. MHC genes are extremely polymorphic and the various allelic products bind a broad spectrum of different peptides, explaining MHC restriction of T-cell recognition. During T-cell development, immature T cells are chosen to mature on the basis of the ability of their clonally expressed TCR to interact weakly with self-MHC molecules expressed in the thymus. During this process of positive selection, self-peptides bound in the MHC groove are also recognized. Thus, the selection of the TCR repertoire is not based solely on the recognition of one or a few self-MHC structures, but on the recognition of those structures modified by the binding of an enormous diversity of self-peptides. Low-affinity interactions between the newly expressed TCRs and self-peptide-MHC molecules select a diverse repertoire that will be fine-tuned to react strongly to pathogen-derived peptides bound to the same MHC in the periphery. Even after maturing and leaving the thymus, T cells continue to depend on survival signals transmitted by the interaction of the TCR with self-peptide-MHC complexes. The long-term homeostasis of naïve T cells depends on continuous tickling of the TCR, and the rebound of the size and complexity of the peripheral lymphocyte pool after any marked loss in cell numbers involves recognition of self-peptide-MHC complexes. Infection often results in massive clonal expansion and development of effector function that initially overrides the balance of lymphocyte pools. When the pathogen is cleared, balance is restored: effector cells are eliminated and a long veteran's retirement is guaranteed for the surviving memory cells, which are no longer dependent on recognition of self-antigen. The veterans may be recalled to action quickly by a subsequent infection with the same pathogen. To provide protection against both unknown new pathogens and the recurrence of old infections, the immune system must maintain separate populations of naïve cells and memory cells. In addition, the system must retain an appropriate mixture of CD4⁺ helper T cells and CD8⁺ cytotoxic T cells in both naïve and memory pools.

Thymocyte development

Bone marrow stem cells enter the thymus and commit to the T lineage in response to signals from the microenvironment⁴. The stages of development of the major lineages of T cells are delineated by the surface expression of the co-receptor molecules CD4 and

CD8. The earliest precursors are CD4⁻CD8⁻ double negative (DN) in phenotype, comprising about 3% of the total thymocyte numbers. These proliferate extensively and become CD4⁺CD8⁺ double positive (DP) cells, which are still immature and are the targets for TCR-based selection events. The MHC molecules expressed by the epithelial elements of the thymus are critically involved in the process of positive selection, which allows a small fraction of the DP population to mature into single positive CD4 lineage or CD8 lineage T cells^{5,6}.

Assembling the TCR

As is the case for antibodies, functional TCR genes are assembled from separate V, D and J region segments by genetic recombination⁷. For T cells expressing $\alpha\beta$ receptors, this process of gene rearrangement occurs in the thymus. Mice and humans carry in their germline a large number of V gene segments, which encode roughly the first 90 residues of a mature TCR α or TCR β chain. The TCR β chain gene is assembled at the DN stage of T-cell development. A short D gene segment is juxtaposed to a short J segment and this is followed by rearrangement of a V segment to the assembled downstream DJ segment. In the case of TCR α chains which rearrange later at the DP stage there are no D segments; a large number of V α segments may rearrange to a large number of J segments. The immunoglobulin-like fold of the TCR chains positions three loops or complementarity determining regions (CDRs) from each chain in close proximity to each other, creating a binding face that will contact antigen. Two of these CDR loops, namely 1 and 2, are encoded by the V gene segment and have only the diversity provided by the number of germline V region gene segments, around 20–70 V α and V β segments in mice and humans. The third CDR loop is created by the juxtaposition of VJ or V(D)J segments and provides much more diversity, a result of the ability of each V segment to rearrange to any (D)J segment compounded by the fact that the joining of the coding sequences is imprecise. Nucleotides may be added or deleted at each of these junctions. If there were only 50 V α and 50 V β genes to encode the TCR repertoire, combinatorial pairing would provide only $50 \times 50 = 2,500$ TCRs. The addition of (D)J segments and the imprecision of joining boosts this number into the billions, with sequence diversity concentrated in the CDR3 loops.

TCR interaction with ligand

Crystal structures of MHC class I and class II molecules have revealed their mode of function (Fig. 1). Although differing in detail, both MHC class I and class II molecules have peptide-binding grooves to present antigen to T cells^{8–10}. In each case, the groove is made up of a 'floor' of β -sheets and 'walls' of long

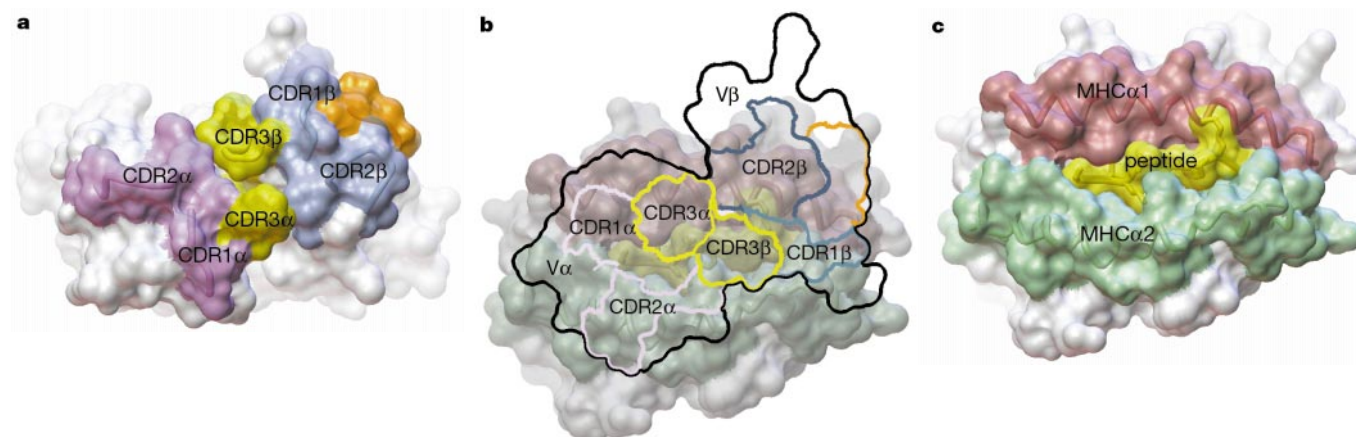


Figure 1 The nature of TCR interaction with peptide–MHC ligand. **a**, Positioning of the three CDR loops of the TCRα and TCRβ chains in the ligand-binding face of the TCR. **b**, Footprint of the TCR on the peptide–MHC complex. **c**, MHC class I molecule with bound

peptide. The long α-helices of MHC enclose the peptide. Reprinted with permission from Garcia *et al.*¹¹. The crystal structure of a TCR interaction with a MHC class II ligand has not been reported but is likely to follow the same rules.

α-helical regions. The most polymorphic residues in the MHC molecules protrude into the groove, thus altering the spectrum of peptides that can be bound. The remarkable feature of MHC–peptide binding is that it is both extremely stable yet promiscuous, allowing many different peptides to bind each allele. Many of the stabilizing interactions are between the backbone of the peptide and the MHC, whereas a few pockets in the groove select for certain side chains on the so-called ‘anchor residues’ of the peptide. MHC class I molecules load peptides soon after synthesis in the endoplasmic reticulum before trafficking to the cell surface. In contrast, the groove of MHC class II molecules is protected from peptide loading until the molecules reach endocytic vesicles. Once loaded with a peptide derived from lysosomal processing, the class-II–peptide complex, like the class-I–peptide complex, is expressed stably at the cell surface.

Mature T cells discriminate between self and foreign by recognition of peptides presented in the groove of self-MHC molecules. The membrane-distal face of the peptide–MHC complex that is presented to the T cell consists of a flat surface made up principally of the α-helices of the MHC surrounding a central peptide. High-resolution structures of the TCR complexed with peptide–MHC ligand have shown how this is achieved (Fig. 1). Crystal structures of a murine TCR interacting with MHC class I and peptide^{11,12}, and two human TCRs complexed with the HLA-A2 class I molecule bound to the same viral peptide^{13,14} have been solved, and all of the structures showed the same basic topology of interaction^{11–16}. The TCR is positioned in a diagonal orientation over the face of the peptide–MHC complex with the CDR1 and CDR2 loops of the TCRα chain lying over the N-terminal half of the peptide and the corresponding regions of the TCRβ chain lying over the C-terminal half. Many of the interactions of the V gene segment-encoded CDR1 and CDR2 regions of the TCR are actually with MHC residues. In particular, the CDR1 and CDR2 loops of the Vα chain have almost identical sites of interactions with peptide–MHC in all structures solved so far, and this match-up may be critical in establishing the overall configuration. The most variable loops of the TCR, namely the CDR3 regions of the TCRα and TCRβ chains, lie centrally and have most contact with protruding peptide side chains. Thus, the most variable region of the TCR (CDR3 loops) has best access to the most variable region of the ligand (peptide).

The unselected TCR repertoire

The lymphoid-specific components of immunoreceptor gene rearrangement, the recombinase-activating gene (RAG) products, are first expressed in thymocytes soon after entry in the thymus and

operate initially on the TCRβ loci. Successful VDJ rearrangement resulting in the production of a functional, in-frame TCRβ chain is tested by assembly of the new gene product (TCRβ) along with an invariant surrogate TCRα chain, pre-Tα, plus signalling components of the TCR, the CD3 complex¹⁷. Proper assembly of this complex with the newly generated TCRβ chain signals the CD4⁺CD8⁺ pre-T-cell to switch off expression of RAG proteins, to go through several rounds of division and to proceed to the CD4⁺CD8⁺ stage of thymocyte maturation. This check point, called β-selection, does not involve the recognition of any ligand by the pre-TCR, as truncated versions of TCRβ and pre-Tα lacking the extracellular domain can trigger progression to the DP compartment¹⁸. The mere assembly of the CD3–pre-Tα–TCRβ complex appears to be sufficient to activate the downstream signalling pathway. It is possible, although perhaps unlikely, that some TCRβ chains are lost in this selection if they cannot assemble with CD3 and/or pre-Tα.

At the DP stage of thymocyte differentiation, the RAG genes are expressed once again, and the cells now focus on rearranging TCR Vα to Jα. Production of a new TCRα chain that can pair with the existing TCRβ chain allows surface expression of the TCRαβ–CD3 complex. Co-expression of randomly rearranged TCRαβ at the DP cell surface is an expression of the germline, or unselected TCR repertoire. The repertoire that is allowed to leave the thymus for the peripheral lymphoid organs is only a subset of this larger pool. To complete maturation, the germline repertoire is positively and negatively selected for interactions with MHC molecules in the thymus. Ultimately, only about 5% of the DPs are allowed to emigrate. In a murine thymus that lacks expression of any MHC molecule, T-cell development is blocked at the DP stage. DP thymocytes from such a MHC-deficient environment are immature and unresponsive to conventional antigen, yet tricks have been used to study the specificity of their germline TCRs. In one case, maturation of the thymocytes was induced in foetal thymic organ culture by receptor engagement with anti-TCR antibodies¹⁹. The mature CD4⁺ T cells that developed were cloned as hybridomas to allow analysis of their TCR specificity. In another case, DP thymocytes from MHC-deficient mice were isolated and reaggregated with MHC wild-type thymus stromal cells. In such reaggregated cultures, the fraction of cells that respond to MHC class I or MHC class II could be scored by examining the upregulation of surface markers that are normally turned on during thymocyte selection²⁰. Both of these approaches showed that the germline TCR repertoire is skewed to the recognition of MHC molecules. Somewhere between 5 and 20% of the DP population from MHC-deficient mice was responsive to the set of MHC molecules expressed in any one mouse

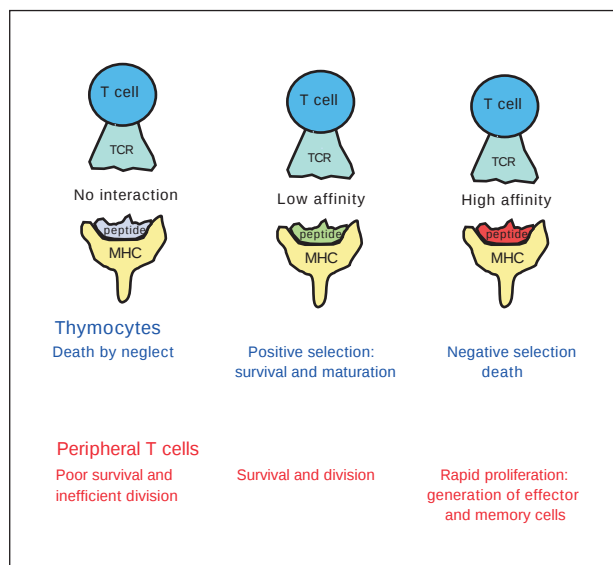


Figure 2 Consequences of TCR interactions with ligands of varying affinity in the thymus and periphery. TCR–MHC interactions may facilitate maturation, survival, proliferation or death. The consequences of TCR binding to peptide–MHC are dictated by the affinity of the interaction, as well as the maturation state of the T cell. Immature thymocytes require a positively selecting signal, which is delivered through low-affinity interactions of their newly rearranged TCR with self-peptide–MHC. If no signal is delivered to the thymocyte because of an inability of the TCR to interact with self-peptide–MHC complexes, it will undergo programmed cell death known as death by neglect. Alternatively, a thymocyte may express a TCR that has high affinity for self-peptide–MHC. These cells must be eliminated by negative selection as they are potentially autoreactive. In the periphery mature T cells require low-affinity interactions with self-peptide–MHC to ensure survival. Furthermore, low-affinity interactions with self-peptide–MHC allow naïve T cells to expand when T-cell numbers are reduced. High-affinity interactions between a naïve T-cell in the periphery and peptide–MHC typically initiate the immune response against a foreign antigen. This leads to rapid proliferation and the development of the effector function culminating in the emergence of a few antigen-experienced cells that join the pool of memory T cells to await a subsequent encounter with antigen.

strain. This is a much higher frequency than would be expected by the generation of receptors through the random rearrangement of gene sequences that do not have an inherent predisposition for interactions with MHC. Evolution has shaped the TCR genes, presumably the CDR1 and CDR2 regions encoded by the V gene segments, so that their $\alpha\beta$ product will preferentially match MHC molecules.

Certain $V\alpha$ gene segments have evolved to interact preferentially with either class I or class II molecules. TCRs bearing these $V\alpha$ segments are preferentially selected into the $CD8^+$ (MHC class I reactive) or $CD4^+$ (MHC class II reactive) T-cell subset during normal thymocyte selection. Swapping individual residues in the CDR1 or CDR2 regions of the $V\alpha$ gene is sufficient to change the selection of these TCRs from the $CD4^+$ to the $CD8^+$ subset²¹. Furthermore, certain mouse strains differ consistently in the ratio of $CD4^+$ to $CD8^+$ T cells in peripheral organs, and one of the loci controlling this phenotype maps with the $TCR\alpha$ locus²².

Positive selection of thymocytes

Even though they express a TCR at the surface, in the absence of MHC recognition $CD4^+CD8^+$ thymocytes will undergo apoptotic death because of receptor neglect (Fig. 2). During their approximate 3-day life span, the DP thymocytes will continue to express the RAG genes and continue to rearrange $TCR\alpha$ chain genes. TCR engagement with MHC class I or class II molecules on the cortical thymic epithelial cells is required to rescue the cell from death by neglect and to switch off further $TCR\alpha$ chain gene rearrangement. Only

those DP thymocytes that bind self-MHC with sufficiently low affinity will be rescued. In contrast, DP cells that express a receptor that interacts with high affinity with MHC in the thymus will be eliminated in a process referred to as negative selection. Almost half of the cells reacting with self-MHC are lost through this mechanism²³ (Fig. 2). The low-affinity recognition of self-MHC not only promotes thymocyte survival but also determines commitment of the DP thymocyte to either the $CD4$ or the $CD8$ lineage^{5,6,24}. Recognition of MHC class I results in commitment to the $CD8$ lineage, whereas recognition of class II results in $CD4^+$ T cells. Thus, function and lineage are correlated with MHC class recognition. How positive selection on MHC class I versus class II is distinguished by the thymocyte is unknown. Lineage choice may be steered by qualitative or quantitative differences in signalling, or may be random but confirmed by later selective events.

The molecular nature of the TCR–MHC recognition event in positive selection has been an area of active research. Initial experiments showed that mutations affecting only the inside of the groove of a murine class I molecule drastically altered the selected repertoire of $CD8^+$ T cells^{25,26}. These experiments suggested strongly that self-peptides bound to the MHC groove are recognized by thymocytes and are critical for positive selection. Further analysis of this recognition required the use of an *in vitro* organ culture system, using thymus lobes from mice deficient in class I expression in which peptides could be added back to reconstitute positive selection. Disruption of the β_2 -microglobulin gene, encoding the light chain of class I molecules, or of the TAP genes encoding a transporter that delivers peptides to nascent class I molecules in the endoplasmic reticulum, leads to a drastic decrease in selection of $CD8$ lineage cells. Surface expression of MHC class I can be restored to some extent by the addition of single peptides or peptide mixtures to organ culture systems. Initial experiments using thymic lobes from mice that express a normal, germline TCR repertoire suggested that a complex mixture of peptides restored selection of $CD8^+$ T cells better than single peptides which bound and stabilized MHC to the same degree^{27,28}. Follow-up experiments in this peptide add-back system using mice that express rearranged TCR transgenes of known ligand specificity suggested a stringent requirement for peptide recognition in the positive selection of $CD8$ lineage T cells^{29–31}. Peptides that bind MHC class I and form a low-affinity ligand for the TCR induced the maturation of functional $CD8^+$ cytotoxic T cells most efficiently³². The measured TCR affinities for foreign peptide–MHC ligands fall into the 1–10 μ M range³³. Compared with this, the affinities for ligands that can efficiently induce positive selection, but cannot induce negative selection, are only 5–9-fold lower³⁴.

Is a single self-peptide actually responsible for the maturation of a cohort of thymocytes? Or do a larger number of self-peptides select each TCR? Natural peptides eluted from MHC class I molecules can induce selection of certain TCR transgenic thymocytes when added back to TAP-deficient thymus lobes^{35,36}, although whether these are expressed at the appropriate level in the thymus cortex is not known. Antibodies that react in a peptide-specific fashion with only a subset of class II molecules prevent the thymic selection of a specific subset of $CD4^+$ T cells³⁷. Certainly, it can be concluded that a subset of self-peptides, and possibly a single self-peptide, bound to self-MHC acts as the ligand that permits the maturation of each T cell. In this way, these self-peptides act as low-affinity mimics or 'stand ins' to prime the immune system to interact later in the periphery with high-affinity ligands for the TCR such as foreign peptide bound to the same MHC³⁸.

In the study of $CD4^+$ T-cell selection in the thymus, researchers have generated mice in which all or a large fraction of the class II molecules in an individual mouse are occupied by the same peptide, and have asked how many T cells can one peptide–MHC ligand select. The invariant chain Ii is a chaperone for MHC class II that contains a region, CLIP, that occupies the class II groove in the

endoplasmic reticulum and prohibits other peptides from binding until the molecules reach endosomal/lysosomal vesicles. At this site, CLIP is removed by the action of the peptide exchange catalyst, H2-M. Mice deficient in H2-M express near normal levels of class-II-peptide on their surface, but in this case, most remain occupied by the CLIP peptide^{39–41}. Another strategy has been to generate a covalently linked class-II-peptide construct that assembles with the single peptide 'sewn in' to the MHC groove and to express these as transgenes in mice that lack wild-type MHC class II^{42,43}.

Initial results from these 'single peptide' systems were remarkable in showing that the mice developed a large and broad CD4⁺ T-cell repertoire. They contained 20–50% of the normal numbers of CD4⁺ T cells, with diverse TCRs as assessed by V β and V α usage. Could a single peptide-MHC complex select a near-normal CD4⁺ T-cell repertoire? More recent analysis has shown quite clearly that the repertoire in these animals is far from normal. First, none of the single-peptide mice can positively select TCRs of known specificity that are selected in normal mice^{44–46}. Second, closer analysis of the TCR α chain usage in the mutant versus wild-type mice shows that the selected repertoire is markedly constrained in the mutants⁴⁷. Last, even when 95% of class II molecules are occupied by a single peptide, most of the CD4⁺ T cells that emerge from the thymus are actually selected on the 5% of molecules that are occupied by a diverse mixture of unknown self-peptides. The CLIP region of Ii was replaced with a tight-binding peptide, so that in mice that were Ii-deficient, 95% of surface MHC class II molecules were occupied by the substituted sequence⁴⁸. Under these conditions, large numbers of CD4⁺ T cells were selected. The critical experiment was to breed these mice to an H2-M-deficient background which resulted in no apparent change in the level of MHC class II expression, but a 70% decline in the number of CD4⁺ T cells selected. The simple conclusion is that the 'background' of diverse peptides bound to class II must have a major role in selection even though such complexes are present at extremely low levels.

Positive selection of immature DP thymocytes by low-affinity interactions with self-peptide-MHC complexes implies that cells at this stage of development are more sensitive to low-affinity ligands than are their mature descendants. This has been shown to be the case, despite the fact that they express 10-fold less TCR on their surface^{49,50}. Low-affinity ligands stimulated early activation events in DP thymocytes but not in mature T cells bearing the same TCR; however, from what we have learned recently about the homeostasis of peripheral T cells, we might assume that under certain circumstances, mature T cells can also respond to low-affinity self-ligands.

Peripheral T-cell homeostasis

After the selection and maturation of T cells in the thymus, naive T cells enter the periphery where they circulate throughout the secondary lymphoid organs awaiting an encounter with antigen. The organization of the many cell types that make up the immune system requires complex homeostatic mechanisms to assure the production and survival of each cell type in appropriate numbers to maintain proper immune function. One can imagine many different mechanisms by which appropriate homeostasis of lymphocyte populations is maintained, including control of the rate of production, division of mature cells, trafficking and cell death.

To maintain homeostasis of lymphocyte populations, the different cell types must be grouped and counted. Peripheral B- and T-cell populations appear to be independently regulated⁵¹. This conclusion is based on the fact that similar numbers of mature B cells are found in the periphery of normal mice and in mice that lack T cells⁵². Likewise, the number of T cells found in normal mice does not increase in B-cell-deficient mice⁵³. In addition, when small numbers of T cells are introduced into depleted hosts, they undergo expansion equivalently in mice deficient in both B and T cells and in mice deficient only in T cells⁵⁴. All these results suggest that independent homeostatic control of T-cell and B-cell numbers.

The pool of peripheral T cells consists of naive T cells and previously activated antigen-experienced memory cells. Interestingly, naive and memory T cells appear to occupy separate homeostatic niches. After reconstitution of T-cell-deficient mice with memory cells, the transferred cells expand to equal the same total number of memory cells in normal mice. Furthermore, it is not possible to increase the size of the memory pool by transferring large numbers of memory T cells, indicating that there is a set point for memory T-cell number even when no naive T cells are present⁵⁵. The separate regulation of memory and naive pools makes teleological sense as it is vital to retain both a diverse naive repertoire that can respond to invasion by yet unknown pathogens and a memory pool that can respond rapidly to previously encountered pathogens. If these two populations were in competition for survival, the diverse repertoire of naive T cells could be replaced by rapidly expanding memory T cells or, conversely, new thymic emigrants could displace memory T cells resulting in the loss of immunological recall⁵¹.

Although CD4⁺ and CD8⁺ cells serve different functions in the immune system and recognize different MHC molecules, many results suggest that they occupy overlapping homeostatic niches⁵¹. First, relatively normal numbers of total T cells are found in mice deficient in MHC class I or II or in the CD8 or CD4 co-receptors, and thus lacking one T-cell subset. Second, small numbers of mature peripheral T cells will expand when introduced into a T-cell-deficient host, and the co-transfer of large numbers of either CD4⁺ or CD8⁺ mature T cells can suppress this expansion^{56,57}. Last, in TCR transgenic mice in which the T-cell population is skewed toward either the CD4⁺ or CD8⁺ subset, total T-cell numbers are typically near normal. Thus, CD4⁺ and CD8⁺ T cells appear to be counted together for purposes of maintaining the homeostasis of T-cell numbers. However, some distinction is clearly made between CD4⁺ and CD8⁺ T cells, as the ratio of CD4⁺ to CD8⁺ T cells is tightly regulated in normal mice and will rapidly readjust after transfer of large numbers of either subset⁵⁶.

The size of the naive pool of T cells remains more or less constant in spite of the continuing contribution of recent thymic emigrants, the loss of T cells due to normal cell death, the expansion and waning of effector populations during immune responses, and the diversion of cells into the memory pool after an immune response. The size of the peripheral T-cell pool does not appear to influence the rate of thymocyte production or export of naive T cells, as changes in the number of peripheral T cells do not change the rate of thymocyte output⁵⁸. These data suggest that homeostatic control over the size of the naive peripheral pool is regulated by the loss of T cells in the periphery⁵⁸. How recent thymic emigrants are incorporated into the peripheral pool is not entirely understood. Clearly, recent thymic emigrants must displace other naive T cells to maintain homeostasis of the peripheral T-cell pool. Newly generated T cells could either randomly displace resident T cells in the periphery or selectively displace resident cells on the basis of a criterion such as age^{51,58,59}. Even in the absence of thymic export, the naive pool remains rather constant in size, indicating that the life span of naive T cells is under homeostatic control^{59,60}.

Relatively little is known about the mechanisms that determine life span, capacity for renewal and survival of naive T cells in the periphery in the absence of antigen. Naive T cells were considered to be quiescent 'spore-like' cells that persist indefinitely without antigenic stimulation^{61–65}; however, evidence now suggests that naive T cells may require input signals to support their survival in the periphery. Expression of transcription factors such as lung Kruppel-like factor and NFAT4 are crucial for the maintenance of a resting, naive population of T cells^{66,67}. In addition, cytokine signals influence the survival, homing and activation of naive T cells, which in turn alter the activation state and homeostasis of the naive compartment^{68,69}. The bcl-2 family of molecules, which are intimately involved in the survival of many cell types, are also involved in the maintenance of the naive pool of T cells⁷⁰. Now it

appears that TCR interactions with self-MHC direct a signal required for maintenance of naïve T cells.

MHC in peripheral maintenance

Although self-peptide-MHC molecules are known to be crucial in the generation of a T-cell repertoire with a wide spectrum of specificities, only recently have data emerged that implicate TCR interaction with self-MHC in the survival and expansion of peripheral T cells (Fig. 2). Several groups have shown that MHC molecules are important in maintaining a peripheral pool of T cells⁷¹.

In many experimental systems, naïve CD4⁺ T cells demonstrate a requirement for MHC class II molecules for long-term survival^{72–75}. CD4⁺ T cells were positively selected in MHC class-II-expressing thymuses and then assayed for survival in a periphery that was engineered to lack expression of the positively selecting MHC class II allele. In all cases, CD4⁺ T cells declined in numbers in the absence of self-MHC class II. One report has shown that expression of MHC class II on dendritic cells is sufficient to support CD4⁺ T cells in the periphery⁷⁴.

An important role for MHC class I in maintaining naïve CD8⁺ T cells has been demonstrated by the transfer of TCR transgenic CD8⁺ T cells into MHC class-I-deficient hosts^{76,77}. Both studies showed that the naïve CD8⁺ T cells fade away in hosts that lack peripheral expression of MHC class I. Furthermore, the MHC class I allele that positively selected the T cell in the thymus is strictly required for naïve T-cell survival⁷⁶. The inability of CD8⁺ T cells to accumulate in mice expressing very low levels of MHC class I when grafted with a wild-type thymus further supports a role for TCR engagement by MHC in the periphery to permit survival of T cells⁷⁸. Although some differences were reported in these studies, the trend remains clear: the retention of both CD8⁺ and CD4⁺ T cells is enhanced when MHC class I and II, respectively, are expressed in the periphery.

MHC in recovery of peripheral homeostasis

The naïve T-cell pool consists mostly of resting cells compared with the memory pool in which cells turn over more rapidly⁶³. Upon close examination, however, low levels of division are observed in naïve T-cell populations in the absence of conventional antigen stimulation^{61,63}. Furthermore, even without conventional antigen stimulation, naïve T cells have a great potential for division when transferred into hosts with low numbers of lymphocytes^{56,60,62,65,79–83}. This proliferation of naïve cells after transfer into lymphopenic hosts has been speculated to involve cytokine release, the expression of environmental antigens or interactions with self-MHC molecules^{57,65,82,84–86}. Recently, the molecular interactions that support the expansion of naïve T cells in lymphopenic environments have become of significant interest, and several groups have studied the function of MHC in this process. There is a clear consensus that both CD4⁺ and CD8⁺ T cells require expression of MHC ligands for efficient homeostatic expansion in lymphopenic hosts^{54,57,77,86–89}.

When polyclonal or TCR transgenic monoclonal CD4⁺ T-cell populations are transferred into syngeneic hosts that have been T-cell depleted, they divide without direct antigenic stimulation. If the CD4⁺ cells are transferred into class II-deficient hosts or hosts expressing MHC class II other than the positively selecting class II molecule, however, little division is observed^{54,57,86,87}. Similarly, both polyclonal and TCR transgenic CD8⁺ T cells expand in wild-type hosts expressing MHC class I, but minimally in hosts with low or no levels of MHC class I molecules^{77,88,89}. Thus, both CD8⁺ and CD4⁺ T cells proliferate much more efficiently in a lymphopenic host expressing MHC class I or MHC class II molecules, respectively. These data show that MHC recognition contributes to division of naïve T cells responding to homeostatic signals.

What is the specificity of the TCR-MHC interaction that promotes homeostasis-driven expansion? Experiments comparing CD4⁺ T cells transferred into hosts with wild-type self-peptide-

MHC class II or hosts expressing MHC class II bound to a limited repertoire of self-peptides show that CD4⁺ T cells undergo expansion efficiently only when transferred into a host with a self-peptide-MHC class II repertoire from the background in which the T cells developed. Division of wild-type CD4⁺ T cells was observed in T-cell-depleted wild-type hosts but only minimally in H2-M-deficient hosts. As expected, CD4⁺ T cells that were positively selected in the H2-M-deficient background were able to divide when transferred into T-cell-depleted H-2M-deficient hosts^{57,86}. These data make a strong case that peptide-specific recognition of MHC class II complexes by T cells promotes division in T-cell-depleted environments^{57,86}. Furthermore, they suggest that this recognition is specific for the same peptides that T cells previously recognized with low affinity during their positive selection in the thymus^{57,86}.

CD8⁺ T cells also exhibit a requirement for specific recognition of peptide-MHC class I complexes to undergo homeostatic expansion. Proliferation of TCR transgenic CD8⁺ T cells in TAP-deficient hosts expressing low levels of MHC class I was restored by the addition of a single MHC class-I-binding peptide⁸⁹. Recognition was peptide specific, as a low-affinity peptide known to positively select the TCR transgenic T cells promoted expansion, whereas a control peptide did not. Thus, the signal for homeostatic proliferation may be characterized by low-affinity TCR interactions with self-peptide-MHC complexes. This suggests that the specific recognition of self-peptide-MHC complexes drives homeostatic proliferation of peripheral T cells, and that this interaction has a specificity similar to the recognition of self-peptide-MHC during positive selection. It is possible that the positively selecting peptide(s) may not necessarily be expressed in the periphery, but that other ligands of similar affinity or avidity for the TCR function in this capacity (Fig. 2).

The response to T-cell depletion

To what extent does homeostasis-driven proliferation of peripheral T cells in a self-peptide-MHC-specific manner contribute to a useful immune system? One possibility is that it ensures the diversity of the peripheral T-cell repertoire. It is appealing to consider that a requirement for peptide-specific recognition during homeostatic proliferation would reinforce the requirement for self-recognition established during positive selection for an individual clone to survive over time and contribute to the long-lived pool of peripheral T cells. Furthermore, the ability of a naïve T cell to receive TCR-self-peptide-MHC signals may determine to what extent a given cell contributes to reconstitution after a decline in T-cell numbers. This could have important implications for peripheral homeostasis after a loss of peripheral T cells in adults, as the ability of the thymus to replenish a naïve T-cell pool with diverse specificities may be insufficient to repopulate the periphery efficiently⁸⁴. The clinical relevancy of such a mechanism certainly exists for patients with T-cell immunodeficiencies resulting from infection (such as AIDS) or from therapeutic treatments (such as chemotherapy or irradiation).

Until recently, it was thought that naïve T cells did not recognize self-peptide-MHC ligands in the periphery. However, the existence of mechanisms for MHC-dependent homeostasis-driven proliferation and survival suggest that low-affinity interactions between the TCR of peripheral T cells and self-peptide-MHC complexes are involved in regulating homeostasis of peripheral T-cell populations. T cells proliferate in response to low-affinity self-ligands when T-cell numbers are reduced, which demonstrates that the sensitivity of the TCR or the response to TCR-mediated signals can be altered by global homeostatic signals. We can envisage at least three different mechanisms by which this could work (Fig. 3). Under normal conditions, T cells crowding through lymph nodes may be competing for ligands on presenting cells. Competition may be for MHC itself or for some other growth factor; lymphopenia would decrease

the competition. Alternatively, T cells may 'count' their own numbers by sending inhibitory signals to each other, or by inhibiting the stimulatory properties of a presenting cell. Decreasing T-cell numbers would sensitize the T cell or activate the presenting cell (Fig. 3).

One implication of naïve T cells being induced to proliferate in T-cell-depleted environments is that what has previously been considered 'maintenance' is likely to be the sum of survival and proliferation. The decline of T cells in the absence of MHC may be partially due to their inability to undergo homeostatic proliferation. Thus, it is necessary at this point to re-evaluate our understanding of T-cell maintenance to accommodate a more complex regulation of T-cell homeostasis.

Maintenance of the memory T-cell pool

Memory T cells often divide in the periphery long after antigen stimulation, even without evidence for the persistence of antigen. Furthermore, the rules for the persistence of memory T cells appear to be different from those that naïve T-cells obey. Recent studies have shown that CD4⁺ and CD8⁺ memory T cells have a much less stringent requirement for TCR–MHC interactions for survival and homeostatic proliferation compared with naïve T cells.

The survival of memory CD8⁺ T cells is much less dependent on MHC class I than that of naïve CD8⁺ T cells. Both polyclonal and TCR transgenic memory CD8⁺ T cells specific for lymphocytic choriomeningitis virus (LCMV) survive indefinitely in MHC-

deficient hosts⁷⁷. Memory CD8⁺ T cells, expressing the TCR transgene specific for the male antigen HY, survived and proliferated without a preference for the restricting allele⁷⁶. However, unlike the LCMV-specific memory T cells, the memory T cells specific for the HY antigen disappeared after transfer into mice expressing low levels of MHC class I^{76,90}. In both experimental systems, CD8⁺ memory T cells were able to proliferate in lymphopenic hosts even when MHC class I was not expressed, showing that homeostatic proliferation of memory CD8⁺ T cells is MHC class I independent^{76,77}. Although some differences in conclusions from these studies may be attributed to distinct experimental systems, the data with polyclonal CD8⁺ T cells specific for many LCMV antigens present a convincing argument for MHC independence for homeostatic proliferation and survival of memory CD8⁺ T cells⁷⁷. Similarly, memory CD4⁺ T cells survive indefinitely in hosts that do not express their positively selecting MHC class II molecule⁹¹ or any class II molecules at all⁹². Thus, CD4⁺ memory T cells appear to have no requirement for the presence of self-MHC class II to support their long-term survival. An obvious corollary of these observations is that neither CD8⁺ nor CD4⁺ memory T cells need continuous exposure to their cognate antigen to survive, a subject of much previous debate.

The response to infection

There is concern that certain infectious agents may overwhelm the immune system because of their high replicative potential; however, the opposite problem may arise. Analyses have shown that, in some cases, the specific immune response can threaten to overwhelm the lymphoid compartment. Use of peptide–MHC tetramers to stain T cells on the basis of TCR specificity, and of single-cell assays to detect interferon- γ secretion, have revolutionized our appreciation of the numerology of the CD8⁺ T-cell response to infection^{93,94}. For mice infected with the natural pathogen LCMV, total CD8⁺ T-cell numbers increase 2–10-fold by eight days after exposure (Fig. 4). Remarkably, most of the CD8⁺ T-cell expansion is antigen-specific! The rare, antigen-specific, naïve CD8⁺ T cells present before infection can divide 13–16 times in this time, so that their frequency may go from 1 in 10⁴–10⁵ to making up most of the CD8⁺ T-cell population. Such rapid proliferation leads to the production of

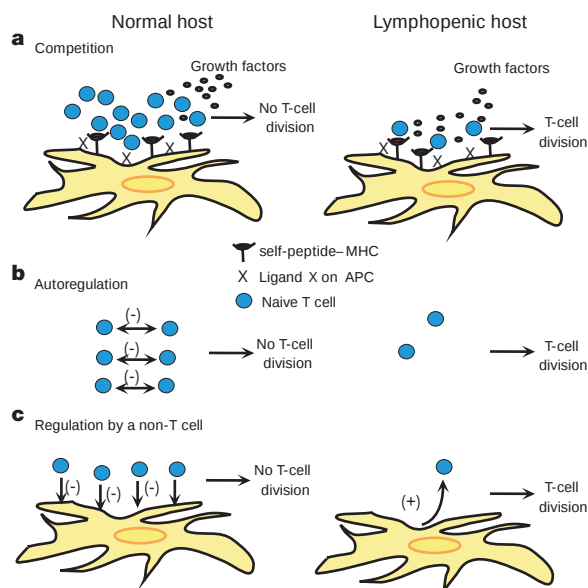


Figure 3 Possible mechanisms to explain the proliferation by naïve T cells in response to low-affinity ligands following T-cell depletion. The lymphopenic state renders naïve T cells capable of responding by proliferation to peptide–MHC complexes that are not stimulatory during normal conditions. **a**, Competition for niches or space on the antigen-presenting cell (APC) would be reduced when T-cell numbers are depleted, thereby allowing greater accessibility to ligands or growth factors. Such decreased competition for non-MHC-derived stimulatory signals such as growth factors or ligands could make signals accessible that would allow T cells to detect self-peptide–MHC complexes resulting in proliferation. It seems improbable that a decrease in direct competition for peptide–MHC complexes explains the ability of T cells to proliferate in response to self-peptide–MHC ligands because CD4⁺ and CD8⁺ T cells recognize distinct ligands yet appear to occupy largely overlapping homeostatic niches. **b**, T cells could somehow 'sense' a drop in T-cell number. T cell/T cell interactions could inhibit proliferation or decrease sensitivity to TCR signals. A release from autoinhibition during lymphopenia would allow T cells to receive a signal for proliferation following interactions with low-affinity ligands for the TCR. **c**, T-cell loss could be detected by another cell type that would upregulate the expression of a factor promoting T-cell division by influencing the T-cell response or sensitivity to TCR interactions with low-affinity peptide–MHC ligands.

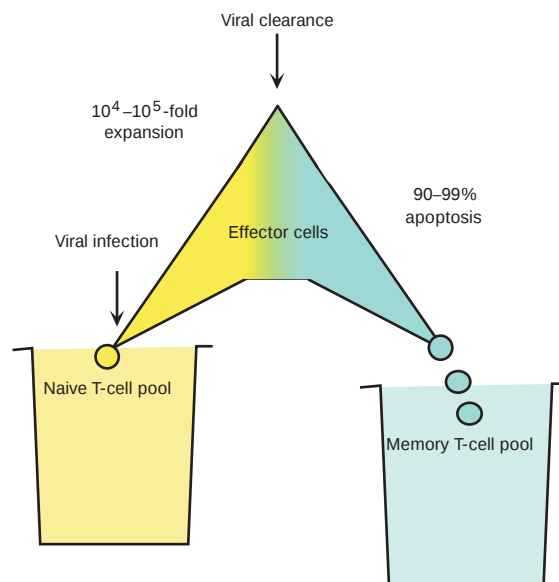


Figure 4 Response to infection and the recovery of homeostasis. The case of the murine CD8⁺ T-cell response to LCMV is depicted. After viral infection, naïve CD8⁺ T cells that recognize LCMV epitopes may go through 13 or more divisions within 8 days, producing massive numbers of effector cells. When antigen is cleared, most of the effector cells die, leaving an expanded population of LCMV-specific T cells, which join the memory pool.

enormous numbers of effectors capable of killing infected cells and releasing cytokines upon antigen encounter. Such a massive CD8⁺ T-cell response is not unique to LCMV infection, as large numbers of antigen-specific T cells are generated in other human and animal infections^{95–99}. In this case of a viral or bacterial pathogen in which multiple pathogen-derived epitopes are known and the specific CD8⁺ T-cell response to each one can be quantitated by tetramer staining, the composition of the memory pool reflects the epitope specificity of the effector population^{94,100}. The dominant epitopes in the initial response remain dominant in the memory repertoire.

The problem for the immune system is how to revert to homeostasis from a situation at the height of response when over half of the CD8⁺ T cells in the infected animal are effector cells specific for one pathogen. The system copes extremely well. During the few days after antigen has been cleared, 90–99% of the effector cells die, leaving an expanded population of new memory cells. The original naïve CD8⁺ T-cell pool is left intact, except for the promotion of the pathogen-specific members, whereas the memory pool adds some new members (Fig. 4). A recent experiment in which memory T cells were irreversibly marked by the activation of a modified granzyme B promoter suggests that a specific subset of effector cells contributes to memory¹⁰¹. What regulates the contraction of the overwhelming numbers of effector CD8⁺ T cells into reasonable numbers of memory cells is unknown. There is no evidence that Fas/FasL or any other tumour-necrosis factor family member is involved in the massive apoptosis of the effector cells¹⁰². Cessation of antigen stimulation and withdrawal of growth factors at the end of the response may lead to death. By whatever mechanism the immune system pulls it off, this represents a remarkable example of the recovery of homeostasis in the T-cell repertoire.

The lives of a T cell

CD4⁺ and CD8⁺ T cells depend on TCR interactions with self-peptides presented by MHC molecules for both their maturation in the thymus and their maintenance in peripheral lymphoid organs. The signals delivered in the periphery through interactions with self maintain the homeostasis of the naïve T-cell pool by keeping cells alive and by stimulating a low level of proliferation when the need arises. Low-affinity interactions with self-peptide–MHC may keep naïve T cells alert and is a way for them to reaffirm their potential to respond to higher affinity interactions with foreign peptide. The recognition of self-peptide–MHC by developing thymocytes does more than simply deliver a survival signal—it also allows the preferential selection of T cells on the basis of their ability to detect self-ligands with high specificity and low affinity. Because the composition of self-peptides bound to MHC is diverse, the specificity of T cells selected to mature is also diverse. The fact that peripheral T cells remain sensitive to self-peptide–MHC suggests that a continuum of self-recognition ensures diversity within the T-cell pool long after positive selection in the thymus is completed.

The factors that govern the production and maintenance of the memory T-cell pool are a mystery. This issue is further compounded by the finding that memory T cells are heterogeneous and can be subclassified into those that retain receptors for entering secondary lymphoid organs and those that preferentially enter inflamed tissues and perform immediate effector function¹⁰³. In contrast to naïve T cells, T cells that have responded to foreign antigen and graduated to the memory pool no longer require self-antigen stimulation to survive. An obvious rationale for this is that once a T cell has responded to a first encounter with a pathogen, it does not need to reaffirm its usefulness in the memory pool. A pathogen that came once may certainly come again. Another reason why memory T cells should be less responsive to self-antigen is that they are potentially more dangerous than naïve cells. Memory T cells are easier to trigger, faster to respond, and may enter tissues more readily than naïve T cells, allowing them to see a broader range of self-antigens. Thus, compared with naïve T cells, they pose an increased threat to

upset the balance of discrimination between self and non-self. On the basis of this realization, one may propose a way in which a fraction of effector cells from a primary response is allowed to revert to memory status. The effector cells that have responded well to foreign antigen, yet are the least sensitive to low-affinity self-peptide–MHC interactions, may be the ones chosen to enter the memory pool. Their lower sensitivity to self-antigen would guard against a harmful autoimmune response due to unforeseen cross-reactivity with self. This lowered sensitivity to self may be based on the specificity of their TCR in a polyclonal population, or on some stochastic change in their ability to perceive low-affinity ligands in a monoclonal setting.

The goal of vaccination is to induce long-lived memory cells. Understanding the developmental progression of T lymphocytes from the thymus, to naïve peripheral cell, to effector cell to memory cell is a challenge for the future. Advances in the way that we are able to identify T cells on the basis of their specificity, and expanding knowledge of signalling pathways will allow rapid progress in this area.

1. Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–710 (1974).
2. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* **317**, 359–361 (1985).
3. Townsend, A. R., Gotch, F. M. & Davey, J. Cytotoxic T cells recognize fragments of the influenza nucleoprotein. *Cell* **42**, 457–467 (1985).
4. Shortman, K. & Wu, L. Early T lymphocyte progenitors. *Annu. Rev. Immunol.* **14**, 29–47 (1996).
5. Jameson, S. C., Hogquist, K. A. & Bevan, M. J. Positive selection of thymocytes. *Annu. Rev. Immunol.* **13**, 93–126 (1995).
6. Kisielow, P. & von Boehmer, H. Development and selection of T cells: facts and puzzles. *Adv. Immunol.* **58**, 87–209 (1995).
7. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
8. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
9. Brown, J. H. *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39 (1993).
10. Scott, C. A., Peterson, P. A., Teyton, L. & Wilson, I. A. Crystal structures of two I-A2-peptide complexes reveal that high affinity can be achieved without large anchor residues. *Immunity* **8**, 319–329 (1998).
11. Garcia, K. C. *et al.* An alpha beta T cell receptor structure at 2.5 Å and its orientation in the TCR–MHC complex. *Science* **274**, 209–219 (1996).
12. Garcia, K. C. *et al.* Structural basis of plasticity in T cell receptor recognition of a self peptide–MHC antigen. *Science* **279**, 1166–1172 (1998).
13. Garboczi, D. N. *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134–141 (1996).
14. Ding, Y. H. *et al.* Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* **8**, 403–411 (1998).
15. Garcia, K. C., Teyton, L. & Wilson, I. A. Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369–397 (1999).
16. Teng, M. K. *et al.* Identification of a common docking topology with substantial variation among different TCR–peptide–MHC complexes. *Curr. Biol.* **8**, 409–412 (1998).
17. Fehling, H. J. & von Boehmer, H. Early alpha beta T cell development in the thymus of normal and genetically altered mice. *Curr. Opin. Immunol.* **9**, 263–275 (1997).
18. Irving, B. A., Alt, F. W. & Killeen, N. Thymocyte development in the absence of pre-T cell receptor extracellular immunoglobulin domains. *Science* **280**, 905–908 (1998).
19. Zerrahn, J., Held, W. & Raulet, D. H. The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636 (1997).
20. Merkenschlager, M. *et al.* How many thymocytes audition for selection? *J. Exp. Med.* **186**, 1149–1158 (1997).
21. Sim, B. C., Zerva, L., Greene, M. I. & Gascoigne, N. R. J. Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science* **273**, 963–966 (1996).
22. Sim, B. C. *et al.* Thymic skewing of the CD4/CD8 ratio maps with the T-cell receptor alphachain locus. *Curr. Biol.* **8**, 701–704 (1998).
23. van Meerwijk, J. P. M. *et al.* Quantitative impact of thymic clonal deletion on the T cell repertoire. *J. Exp. Med.* **185**, 377–383 (1997).
24. Robey, E. & Fowlkes, B. J. Selective events in T cell development. *Annu. Rev. Immunol.* **12**, 675–705 (1994).
25. Nikolic-Zugic, J. & Bevan, M. J. Role of self-peptides in positively selecting the T-cell repertoire. *Nature* **344**, 65–67 (1990).
26. Sha, W. C. *et al.* Positive selection of transgenic receptor-bearing thymocytes by Kb antigen is altered by Kb mutations that involve peptide binding. *Proc. Natl Acad. Sci.* **87**, 6186–6190 (1990).
27. Hogquist, K. A., Gavin, M. A. & Bevan, M. J. Positive selection of CD8⁺ T cells induced by major histocompatibility complex binding peptides in fetal thymic organ culture. *J. Exp. Med.* **177**, 1469–1473 (1993).
28. Ashton-Rickardt, P. G., Van Kaer, L., Schumacher, T. N., Ploegh, H. L. & Tonegawa, S. Peptide contributes to the specificity of positive selection of CD8⁺ T cells in the thymus. *Cell* **73**, 1041–1049 (1993).
29. Ashton-Rickardt, P. G. *et al.* Evidence for a differential avidity model of T cell selection in the thymus. *Cell* **76**, 651–663 (1994).

30. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
31. Sebzda, E. *et al.* Mature T cell reactivity altered by peptide agonist that induces positive selection. *J. Exp. Med.* **183**, 1093–1104 (1996).
32. Hogquist, K. A., Jameson, S. C. & Bevan, M. J. Strong agonist ligands for the T cell receptor do not mediate positive selection of functional CD8⁺ T cells. *Immunity* **3**, 79–86 (1995).
33. Davis, M. M. *et al.* Ligand recognition by alpha beta T cell receptors. *Annu. Rev. Immunol.* **16**, 523–544 (1998).
34. Alam, S. M. *et al.* T-cell-receptor affinity and thymocyte positive selection. *Nature* **381**, 558–559 (1996).
35. Hogquist, K. A. *et al.* Specific recognition of thymic self-peptides induces the positive selection of cytotoxic T lymphocytes. *Immunity* **7**, 221–231 (1997).
36. Hu, Q. *et al.* Specific recognition of thymic self-peptides induces the positive selection of cytotoxic T lymphocytes. *Immunity* **7**, 221–231 (1997).
37. Baldwin, K. K., Reay, P. A., Wu, L., Farr, A. & Davis, M. M. A T cell receptor-specific blockade of positive selection. *J. Exp. Med.* **189**, 13–24 (1999).
38. Kersh, G. J. & Allen, P. M. Essential flexibility in the T-cell recognition of antigen. *Nature* **380**, 495–498 (1996).
39. Fung-Leung, W. P. *et al.* Antigen presentation and T cell development in H2-M-deficient mice. *Science* **271**, 1278–1281 (1996).
40. Martin, W. D. *et al.* H2-M mutant mice are defective in the peptide loading of class II molecules, antigen presentation, and T cell repertoire selection. *Cell* **84**, 543–550 (1996).
41. Miyazaki, T. *et al.* Mice lacking H2-M complexes, enigmatic elements of the MHC class II peptide-loading pathway. *Cell* **84**, 531–541 (1996).
42. Ignatowicz, L., Kappler, J. & Marrack, P. The repertoire of T cells shaped by a single MHC/peptide ligand. *Cell* **84**, 521–541 (1996).
43. Fukui, Y. *et al.* Positive and negative CD4⁺ thymocyte selection by a single MHC class II/peptide ligand affected by its expression level in the thymus. *Immunity* **6**, 401–410 (1997).
44. Grubin, C. E., Kovats, S., deRoos, P. & Rudensky, A. Y. Deficient positive selection of CD4⁺ T cells in mice displaying altered repertoires of MHC class II-bound self-peptides. *Immunity* **7**, 197–208 (1997).
45. Tourne, S. *et al.* Selection of a broad repertoire of CD4⁺ T cells in H-2Ma0/0 mice. *Immunity* **7**, 187–195 (1997).
46. Surh, C. D., Lee, D. S., Fung-Leung, W. P., Karlsson, L. & Sprent, J. Thymic selection by a single MHC/peptide ligand produces a semidiverse repertoire of CD4⁺ T cells. *Immunity* **7**, 209–219 (1997).
47. Sant'Angelo, D. B. *et al.* The imprint of intrathymic self-peptides on the mature T cell receptor repertoire. *Immunity* **7**, 517–524 (1997).
48. Barton, G. M. & Rudensky, A. Y. Requirement for diverse, low-abundance peptides in positive selection of T cells. *Science* **283**, 67–70 (1999).
49. Davey, G. M. *et al.* Preselection thymocytes are most sensitive to T cell receptor stimulation than mature T cells. *J. Exp. Med.* **188**, 1867–1874 (1998).
50. Lucas, B., Stefanova, I., Yasutomo, K., Dautigny, N. & Germain, R. N. Divergent changes in the sensitivity of maturing T cells to structurally related ligands underlies formation of a useful T cell repertoire. *Immunity* **10**, 367–376 (1999).
51. Tanchot, C., Rosado, M. M., Agnes, F., Freitas, A. A. & Rocha, B. Lymphocyte homeostasis. *Semin. Immunol.* **9**, 331–337 (1997).
52. Mombaerts, P. *et al.* Mutations in T-cell antigen receptor genes alpha and beta block thymocyte development at different stages. *Nature* **360**, 225–231 (1992).
53. Kitamura, D., Roes, J., Kuhn, R. & Rajewsky, K. A B-cell deficient mouse by targeted disruption of the membrane exon of the immunoglobulin mu chain gene. *Nature* **350**, 423–426 (1991).
54. Bender, J., Mitchell, T., Kappler, J. & Marrack, P. CD4⁺ T cell division in irradiated mice requires peptides distinct from those responsible for thymic selection. *J. Exp. Med.* **190**, 367–374 (1999).
55. Tanchot, C. & Rocha, B. The peripheral T cell repertoire: independent homeostatic regulation of virgin and activated CD8⁺ T cell pools. *Eur. J. Immunol.* **25**, 2127–2136 (1995).
56. Rocha, B., Dautigny, N. & Pereira, P. Peripheral T lymphocytes: expansion potential and homeostatic regulation of pool sizes and CD4/CD8 ratios *in vivo*. *Eur. J. Immunol.* **19**, 905–911 (1989).
57. Ernst, B., Lee, D. Chang, J. M., Sprent, J. & Surh, C. D. The peptide ligands mediating positive selection in the thymus control T cell survival and homeostatic proliferation in the periphery. *Immunity* **11**, 173–181 (1999).
58. Berzins, S. P., Boyd, R. L. & Miller, J. F. A. P. The role of the thymus and recent thymic migrants in the maintenance of the adult peripheral lymphocyte pool. *J. Exp. Med.* **187**, 1839–1848 (1998).
59. Tanchot, C. & Rocha, B. Peripheral selection of T cell repertoires: the role of continuous thymus output. *J. Exp. Med.* **186**, 1099–1106 (1997).
60. Bell, E. B., Sparshott, S. M., Drayson, M. T. & Ford, W. L. The stable and permanent expansion of functional T lymphocytes in athymic nude rats after a single injection of mature T cells. *J. Immunol.* **139**, 1379–1384 (1987).
61. Sprent, J. Lifespans of naive, memory and effector lymphocytes. *Curr. Opin. Immunol.* **5**, 433–438 (1993).
62. Sprent, J., Schaefer, M., Hurd, M., Surh, C. D. & Ron, Y. Mature murine B and T cells transferred to SCID mice can survive indefinitely and many maintain a virgin phenotype. *J. Exp. Med.* **174**, 717–728 (1991).
63. Tough, D. F. & Sprent, J. Turnover of naive- and memory-phenotype T cells. *J. Exp. Med.* **179**, 1127–1135 (1994).
64. von Boehmer, H. & Hafen, K. The life span of naive α/β T cells in secondary lymphoid organs. *J. Exp. Med.* **177**, 891–896 (1993).
65. Bruno, L., von Boehmer, H. & Kirberg, J. Cell division in the compartment of naive and memory T lymphocytes. *Eur. J. Immunol.* **26**, 3179–3184 (1996).
66. Kuo, C. T., Veselitis, M. L. & Leiden, J. M. LKLF: a transcriptional regulator of single-positive T cell quiescence and survival. *Science* **277**, 1986–1990 (1997).
67. Oukka, M. *et al.* The transcription factor NFAT4 is involved in the generation and survival of T cells. *Immunity* **9**, 295–304 (1998).
68. Lodolce, J. P. *et al.* IL-15 receptor maintains lymphoid homeostasis by supporting lymphocyte homing and proliferation. *Immunity* **9**, 669–676 (1998).
69. Nakajima, H., Shores, E. W., Noguchi, M. & Leonard, W. J. The common cytokine receptor γ chain plays an essential role in regulating lymphoid homeostasis. *J. Exp. Med.* **185**, 189–195 (1997).
70. Veis, D. J., Sorenson, C. M., Shutter, J. R. & Korsmeyer, S. J. Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair. *Cell* **75**, 229–240 (1993).
71. Freitas, A. A. & Rocha, B. Peripheral T cell survival. *Curr. Opin. Immunol.* **11**, 152–156 (1999).
72. Kirber, J., Berns, A. & von Boehmer, H. Peripheral T cell survival requires continual ligation of the T cell receptor to major histocompatibility complex-encoded molecules. *J. Exp. Med.* **186**, 1269–1275 (1997).
73. Takeda, S., Rodewald, H.-R., Arakawa, H., Bluethmann, H. & Shimizu, T. MHC class II molecules are not required for survival of newly generated CD4⁺ T cells, but affect their long-term life span. *Immunity* **5**, 217–228 (1996).
74. Brocker, T. Survival of mature CD4 T lymphocytes is dependent on major histocompatibility complex class II-expressing dendritic cells. *J. Exp. Med.* **186**, 1223–1232 (1997).
75. Rooke, R., Waltzinger, C., Benoist, C. & Mathis, D. Targeted complementation of MHC class II deficiency by intrathymic delivery of recombinant adenoviruses. *Immunity* **7**, 123–134 (1997).
76. Tanchot, C., Lemonnier, F. A., Perarnau, B., Freitas, A. A. & Rocha, B. Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science* **276**, 2057–2062 (1997).
77. Murali-Krishna, K. *et al.* Persistence of memory CD8 T cells in MHC class I deficient mice. *Science* (in the press).
78. Nesic, D. & Vukmanovic, S. MHC class I is required for peripheral accumulation of CD8⁺ thymic emigrants. *J. Immunol.* **160**, 3705–3712 (1998).
79. Bell, E. B., Sparshott, S. M., Drayson, M. T. & Hunt, S. V. The origin of T cells in permanently reconstituted old athymic nude rats. Analysis using chromosome or allotype markers. *Immunology* **68**, 547–556 (1989).
80. McDonagh, M. & Bell, E. B. The survival and turnover of mature and immature CD8 T cells. *Immunology* **84**, 514–520 (1995).
81. Pereira, P. & Rocha, B. Post-thymic *in vivo* expansion of mature $\alpha\beta$ T cells. *Int. Immunol.* **3**, 1077–1080 (1991).
82. Oehen, S. & Brduscha-Riem, K. Naive cytotoxic T lymphocytes spontaneously acquire effector function in lymphocytopenic recipients: A pitfall for T cell memory studies? *Eur. J. Immunol.* **29**, 608–614 (1999).
83. Sprent, J., Surh, C. D. & Tough, D. Fate of T and B cells transferred to SCID mice. *Res. Immunol.* **145**, 328–331 (1994).
84. Bell, E. B. & Sparshott, S. M. The peripheral T-cell pool: regulation by non-antigen induced proliferation? *Semin. Immunol.* **9**, 347–353 (1997).
85. Mackall, C. L. *et al.* Thymic-independent T cell regeneration occurs via antigen-driven expansion of peripheral T cells resulting in a repertoire that is limited in diversity and prone to skewing. *J. Immunol.* **156**, 4609–4616 (1996).
86. Viret, C., Wong, F. S. & Janeway, C. S. Designing and maintaining the mature TCR repertoire: the continuum of self-peptide:self-MHC complex recognition. *Immunity* **10**, 559–568 (1999).
87. Beutner, U. & MacDonald, H. R. TCR-MHC class II interaction is required for peripheral expansion of CD4 cells in a T cell-deficient host. *Int. Immunol.* **10**, 305–310 (1998).
88. Kieper, W. C. & Jameson, S. C. Homeostatic expansion and phenotypic conversion of naive T cells in response to self peptide/MHC ligands. *Proc. Natl. Acad. Sci. USA* (in the press).
89. Goldrath, A. W. & Bevan, M. J. Low affinity ligands for the TCR drive proliferation of mature CD8⁺ T cells in lymphopenic hosts. *Immunity* **11**, 183–190 (1999).
90. Markiewicz, M. A. *et al.* Long-term T cells memory requires the surface expression of self-peptide/major histocompatibility complex molecules. *Proc. Natl. Acad. Sci. USA* **95**, 3065–3070 (1998).
91. Garcia, S., DiSanto, J. & Stockinger, B. Following the development of a CD4 T cell immune response *in vivo*: from activation to memory formation. *Immunity* **11**, 163–171 (1999).
92. Swain, S., Hui, H. & Huston, G. Class II-independent generation of CD4 memory T cells from effectors. *Science* (in the press).
93. Butz, E. A. & Bevan, M. J. Massive expansion of antigen-specific CD8⁺ T cells during an acute virus infection. *Immunity* **8**, 167–175 (1998).
94. Murali-Krishna, K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
95. Busch, D. H., Pilip, I. M., Vijh, S. & Pamer, E. G. Coordinate regulation of complex T cell populations responding to bacterial infection. *Immunity* **8**, 353–362 (1998).
96. Callan, M. F. *et al.* Direct visualization of antigen-specific CD8⁺ T cells during the primary immune response to Epstein-Barr virus *in vivo*. *J. Exp. Med.* **187**, 1395–1402 (1998).
97. Cose, S. C., Jones, C., Wallace, M. E., Heath, W. R. & Carbone, F. R. Antigen-specific CD8⁺ T cell subset distribution in lymph nodes draining the site of herpes simplex virus infection. *Eur. J. Immunol.* **27**, 2310–2316 (1997).
98. Flynn, K. J. *et al.* Virus-specific CD8⁺ T cells in primary and secondary influenza pneumonia. *Immunity* **8**, 683–691 (1998).
99. Zimmermann, C. & Pircher, H. A novel approach to visualize polyclonal virus-specific CD8 T cells *in vivo*. *J. Immunol.* **162**, 5178–5182 (1999).
100. Busch, D. H., Pilip, I. & Pamer, E. G. Evolution of a complex T cell receptor repertoire during primary and recall bacterial infection. *J. Exp. Med.* **188**, 61–70 (1998).
101. Jacob, J. & Baltimore, D. Modelling T-cell memory by genetic marking of memory T cells *in vivo*. *Nature* **399**, 593–597 (1999).
102. Razvi, E. S., Jiang, Z., Woda, B. A. & Welsh, R. M. Lymphocyte apoptosis during the silencing of the immune response to acute viral infections in normal, Ipr, and Bcl-2-transgenic mice. *Am. J. Pathol.* **147**, 79–91 (1995).
103. Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).

Correspondence should be addressed to M.J.B. (e-mail: mbevan@u.washington.edu).

Structural changes linked to proton translocation by subunit *c* of the ATP synthase

Vinit K. Rastogi & Mark E. Girvin

Biochemistry Department, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

F₁F₀ ATP synthases use a transmembrane proton gradient to drive the synthesis of cellular ATP. The structure of the cytosolic F₁ portion of the enzyme and the basic mechanism of ATP hydrolysis by F₁ are now well established, but how proton translocation through the transmembrane F₀ portion drives these catalytic changes is less clear. Here we describe the structural changes in the proton-translocating F₀ subunit *c* that are induced by deprotonating the specific aspartic acid involved in proton transport. Conformational changes between the protonated and deprotonated forms of subunit *c* provide the structural basis for an explicit mechanism to explain coupling of proton translocation by F₀ to the rotation of subunits within the core of F₁. Rotation of these subunits within F₁ causes the catalytic conformational changes in the active sites of F₁ that result in ATP synthesis.

The F₁F₀ ATP synthase is the main source of cellular ATP. It uses a transmembrane proton gradient to drive the synthesis of ATP from ADP and phosphate. The ATP synthase (Fig. 1) consists of a water-soluble F₁ portion, whose crystal structure has been solved^{1,2}, and a transmembrane F₀ portion, for which little structural information exists³. Proton transport through F₀ drives the release of ATP product on F₁ by long-range conformational changes. F₁ consists of five subunit types in an $\alpha_3\beta_3\gamma_1\delta_1\epsilon_1$ stoichiometry, with a ring of α - and β -subunits alternating around a single γ -subunit¹. Differential interactions of the three β -subunits with the single γ -subunit induce asymmetry at the three catalytic sites. The catalytic sites on each of the three β -subunits cycle between three binding states for substrates and products in an 'alternating sites' mechanism^{3,4}. ATP hydrolysis has been shown to drive rotation of the γ -subunit within the core of F₁ (refs 5–7). In the synthetic direction, proton translocation through F₀ must drive rotation of the γ -subunit within F₁.

Three types of subunit make up F₀, in an $a_1b_2c_{12}$ stoichiometry^{8,9}. Low-resolution images and biochemical data indicate an annular arrangement of the twelve *c* subunits, with subunits *a* and *b*₂ on the periphery of the cylinder^{10–14}. The long helical subunit-*b* dimer links subunit *a* of F₀ with the δ -subunit and one of the α -subunits of F₁ (refs 15–17). The ring of *c* subunits makes contact with subunit *a* in F₀ (ref. 14). The subunit *c* ring is also linked to the γ -subunit of the F₁ core both directly¹⁸, and indirectly through mutual contacts with the ϵ -subunit¹⁹. An essential carboxylate, Asp 61 in *Escherichia coli* subunit *c*²⁰, translocates protons at the interface between subunits *a* and *c*. Subunit *c* of *E. coli* is a 79-residue protein which folds as two transmembrane segments connected by a polar loop^{21–23}. Conserved residues in the loop form part of the physical link between subunit *c* of F₀ and the $\gamma\epsilon$ -subunits of the F₁ portion of the enzyme^{18,19}. Current models postulate that protonation and subsequent ionization of the Asp 61 side chain in subunit *c* lead to rotation of the *c*₁₂ oligomer with respect to the *a* and *b*₂ subunits²⁴. Rotation of the *c*₁₂ oligomer in turn causes rotation of the ϵ - and γ -subunits with respect to the catalytic sites on the β -subunits in F₁, through interactions between subunits *c* and $\epsilon\gamma$. The structure of the Asp 61-protonated form of subunit *c* at pH 5 and the predicted packing between subunits in the *c*₁₂ oligomer have been reported²¹. Here we describe the structural changes within the protein induced by deprotonation of Asp 61 at pH 8. The structures of the two forms of the protein, together with the known interactions between subunit *c* monomers and between subunits *a* and *c*, lead us to

propose a mechanism for the rotation within F₀ that drives ATP synthesis on F₁.

Structure of subunit *c* with Asp 61 deprotonated

The three-dimensional structure of subunit *c* at pH 8 was defined by distance constraints derived from ¹³C- and ¹⁵N-resolved three-dimensional nuclear Overhauser enhancement spectroscopy (3D NOESY) data²⁵. The purification, sample preparation, and ¹H, ¹³C, and ¹⁵N resonance assignments of subunit *c* at pH 8, where Asp 61 is deprotonated ($pK_a = 7.1$ (ref. 26)), have been described²⁷. The protein was well behaved at pH 8, and the nuclear magnetic resonance (NMR) data were of uniformly high quality. Representative NOESY data in Fig. 2a show the ¹³C-edited nuclear Overhauser enhancements (NOEs) from the H ^{α} of residues in the polar loop. Different sets of long-range NOEs were observed for the protein at pH 8 and pH 5 (Fig. 2b). Initial structures of subunit *c* were calculated by high-temperature simulated annealing using torsion angle dynamics²⁸ within the CNS program²⁹. During refinement, hydrogen-bonding constraints were used for slowly exchanging amide protons that had reduced temperature coefficients for the amide proton chemical shifts³⁰. The final stages of refinement also

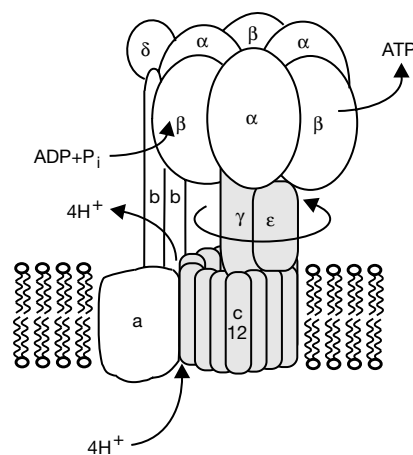


Figure 1 Schematic diagram of F₁F₀ ATP synthase. Subunits shown in white are thought to remain fixed with respect to each other during proton translocation and ATP synthesis. Subunits shown in grey are thought to rotate as a unit with respect to the fixed subunits during catalysis.

included ^{13}C and ^1H chemical shift constraints^{31,32}. Table 1 lists the constraints used for the structure calculations and the statistics for the final structural models.

The structural models fit the NMR data well, with no distance constraint violations greater than 0.4 Å. The positions of all backbone and most side-chain atoms were well defined by the NMR constraints. The models shown in Fig. 3a were superimposed using the backbone coordinates for residues Tyr 10–Phe 35 and Arg 50–Val 74. For these stretches, the root mean squared deviations (r.m.s.d.s) for the backbone atoms and for all non-hydrogen atoms were 0.5 Å and 1.0 Å, respectively. The structure consisted of a curved amino-terminal helix extending from Glu 2 through Arg 41, connected by a short structured loop to the carboxy-terminal helix. The C-terminal helix extended from Ile 46 through Val 74, and changed direction at Pro 64, just below Asp 61.

Conformational changes on deprotonation of Asp 61

The structures of subunit *c* in both protonation states share a similar

helix–loop–helix topology (Fig. 3b). Comparison of the two structures revealed that the main conformational change following deprotonation of Asp 61 was a dramatic rotation of the C-terminal helix with respect to the N-terminal helix, as can be seen by comparing the positions of labelled side chains in Fig. 3b. The N-terminal helix was virtually unaffected, with backbone r.m.s.d.s of 0.8 Å between protonated and deprotonated forms (superimposed using only the backbone atoms of Val 15–Arg 41 in Fig. 3c). The local geometries of the C-terminal helices were also similar in the two protonation states, with backbone atom r.m.s.d.s of 1.1 Å (superimposed using only the backbone atoms of Ile 55–Met 75 in Fig. 3c). However, in the deprotonated form, the C-terminal helix as a unit rotated about its axis by 140°. This rotation induced by deprotonating Asp 61 led to conformational changes in the short structured loop, particularly for residues Gln 42–Asp 44 (Fig. 3b). These residues make up most of the binding interface for the ϵ -subunit of F_1 (ref. 19).

Independent disulphide cross-linking data confirm that this reorientation of the C-terminal helix occurs in the intact enzyme. A set of residues was identified in the C-terminal helix of subunit *c* which, when mutated to cysteine, could be cross-linked to a corresponding set of cysteine residues introduced in the penultimate helix of subunit *a* (ref. 14). These residues within subunit *c* are buried in the packing interface between the N- and C-terminal helices in the structure of the protonated monomer, but they form an exposed helical face in the deprotonated structure reported here (see Fig. 4a). This rotation of the C-terminal helix as the Asp 61 side chain cycles between its two protonation states provides the first structural insight into the mechanism by which H^+ translocation can be linked to rotations within F_0 .

Model of the c_{12} oligomer

The relationship between structural changes in subunit *c*, H^+ translocation through F_0 and rotational coupling to F_1 can be visualized better in a model of the subunit *c* oligomer. Although the high-resolution structure of such an entity does not yet exist, it was possible to combine the NMR structures described here with distance constraints derived from disulphide cross links between subunit *c* monomers¹³ to calculate a family of models for the c_{12} oligomer. The successful synthesis of these two independent types of data increases confidence in the biological relevance of the two monomeric NMR structures, and demonstrates the value of determining structures of membrane protein subunit monomers.

For the calculations, we assumed that there are twelve contiguous copies of subunit *c* per complex^{9,10,13}, with one subunit *c* in the Asp 61 deprotonated form and the remaining eleven in the protonated form. The appropriate set of NOE-derived distance constraints were included for each monomer. Inter-subunit C^α – C^α upper

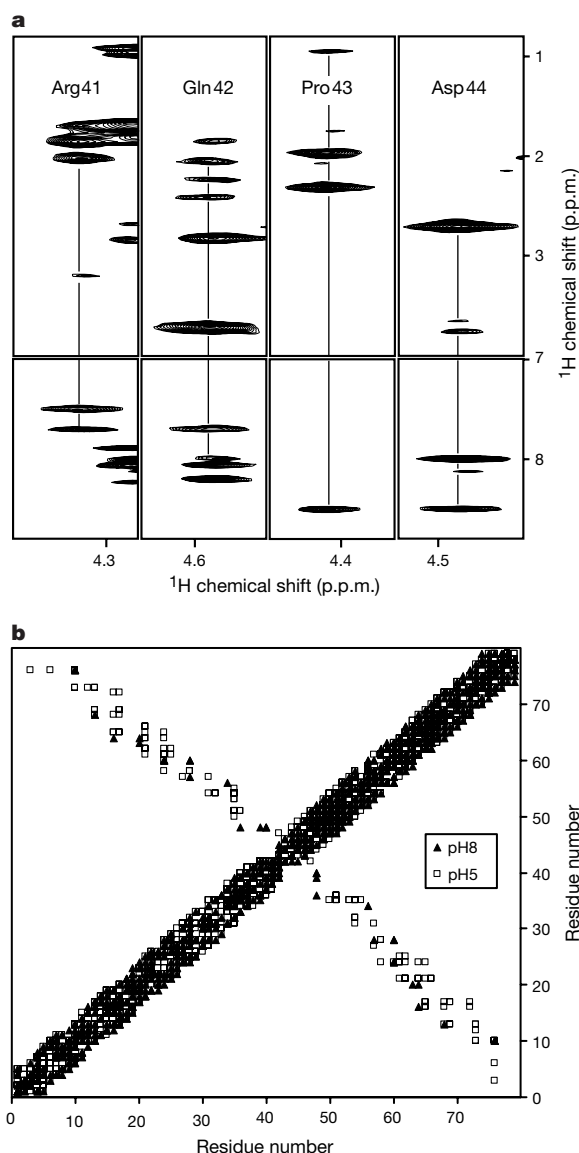


Figure 2 Portions of a 3D $^1\text{H}/^{13}\text{C}$ NOESY–heteronuclear single quantum coherence (HSQC) spectrum of subunit *c* at pH 8, and summary of observed NOEs. **a**, Series of slices from a ^{13}C -edited NOESY–HSQC spectrum (80 ms mixing time) showing NOEs from the H^α of residues in the polar loop. **b**, Comparison of inter-residue NOEs for subunit *c* shows that different sets of interhelical contacts are observed at pH 8 (filled triangles) and pH 5 (open squares).

Table 1 Structural statistics for subunit <i>c</i> at pH 8	
NOE distance restraints (total 703)	
intra-residue	203
sequential	214
medium	267
long range (inter-helix)	19
Backbone hydrogen bonds	56
^{13}C chemical shifts	144
^1H chemical shifts	288
r.m.s.d. of atomic coordinates (residues 10–35, 50–74)	
backbone (Å)	0.50
all non-hydrogens (Å)	1.00
Deviations from experimental distance restraints	
largest (Å)	0.34
r.m.s. (Å)	0.025
Ramachandran plot	
residues in most favoured regions	94%
residues in additional allowed regions	6%
r.m.s.d. from ideal geometry	
bond lengths (Å)	0.003
bond angles (deg)	0.507
impropers (deg)	0.399

distance constraints of 12 Å were used for the 12 cysteine cross-links of ref. 13 for which the mutants grew robustly, and high-yield multimeric products were formed: inter-subunit residue pairs 8 → 8, 11 → 11, 15 → 15, 26 → 26, 30 → 30, 16 → 14, 68 → 13, 20 → 62, 20 → 66, 58 → 28, 65 → 21 and 72 → 14. Because of the large change in orientation of the C-terminal helix that follows deprotonation of Asp 61 and the lack of c_{12} cross-linked oligomers involving only C-terminal residues¹³, no inter-subunit constraints were imposed on this region of the deprotonated form of subunit *c*. A linear array of 12 *c* subunits separated by 25 Å was created to eliminate any bias from the starting configuration, and a family of 20 structural models was calculated using the torsion angle dynamics–simulated annealing protocols^{28,29} typically used for NMR structure determinations. Non-crystallographic symmetry constraints were added for the 11 protonated monomers and a second cycle of simulated annealing was performed. In the best 10 models, no NOE or disulphide-derived distance constraints were violated by more than 0.4 Å, showing that models could be calculated for the oligomer that were entirely consistent with both the constraints determined for the monomer in solution and the inter-molecular constraints determined in the intact F_0 complex. These constraints defined a unique set of similar conformations, with r.m.s.d.s between all backbone atoms of only 0.6 Å. The structures of the monomers within the model for the complex were also extremely similar to the corresponding NMR structures. The r.m.s.d.s between the backbone atoms in the transmembrane segments of the NMR structures and those in the modelled oligomer were 0.6 Å for the protonated form and 0.8 Å for the deprotonated form. For clarity, only a single member of the family of oligomer models is shown in Fig. 4a.

In the oligomeric model, the *c* subunits were arranged in a ring, with their N termini towards the centre. The inner and outer diameters of the annular structure were 25 and 62 Å, consistent with electron spectroscopic imaging¹⁰ and atomic force microscopy^{11,12}. Each Asp 61-protonated monomer formed the predicted packing interface with its neighbours. The packing surfaces were complementary in shape and charge, and each interface buried 2,240 Å² of surface area. The N-terminal helices packed more closely (11 Å centre-to-centre) than the C-terminal helices (14.5 Å centre-to-centre), permitting some access to Asp 61 from the outer face of the ring. Within the monomers, the N- and C-helices were 9.5 Å apart, and the N- and C-helices of adjacent monomers were 14 Å apart. The Asp 61-deprotonated monomer

also packed well with its protonated neighbours, its N-terminal helix forming the same intermolecular contacts as the protonated form, and the rotated C-terminal helix forming a new set of favourable contacts with its neighbours.

Structural models for an oligomer of 12 Asp 61-protonated *c* subunits have been proposed^{13,34}. Restrained molecular mechanics and dynamics were used³³ to model an oligomer consistent with disulphide cross links¹³ and the secondary structure determined by NMR²¹. The modelled oligomer was generally similar to that derived here; but, as no interhelical constraints within monomers were included, the orientation of the C-terminal helix with respect to the N-terminal helix within the individual monomers differed from both the model reported here and the solution NMR structure. The C-terminal helix was offset by up to 8 Å from the N-terminal helix around the circumference of the ring, and the backbone atom positions of the monomers differed by 2.6 Å r.m.s.d. from the refined NMR structure of the Asp 61-protonated subunit. An earlier model of the c_{12} oligomer had been built³⁴ based on a partial NMR structure³⁵ and sequence analysis. In this model the N and C termini were reversed, with the N-terminal helix facing the outside of the ring, which does not appear to be consistent with more recent disulphide cross-linking results^{13,14}.

The deprotonated Asp 61 side chain should lie at the active site for proton translocation in the c_{12} oligomer. Details of this region are shown in Fig. 4b. The deprotonated Asp 61 side chain faced the Ala 24 and Ile 28 side chains in its N-terminal helix. Considerable experimental evidence supports such a location⁸. Even more interesting was the juxtaposition of the deprotonated Asp 61 of one monomer with the protonated Asp 61 of the adjacent monomer, forming an obvious site for interaction with Arg 210 in subunit *a*. This arginine is essential for the H⁺ translocation that is coupled to ATP synthesis³⁶, and has been predicted to assist in the protonation cycle of Asp 61 in subunit *c*.

Model of subunit *a* and the ac_{12} complex

To derive an explicit mechanism for proton translocation, models for subunit *a* and its interaction with the c_{12} oligomer were needed. We generated a family of models for the four consensus transmembrane helices of subunit *a*^{37–40} by standard NMR structure calculation methods using only published biochemical data as constraints. The constraints used were: helical backbone torsion angles and hydrogen-bonding constraints for residues 100–124, 138–166,

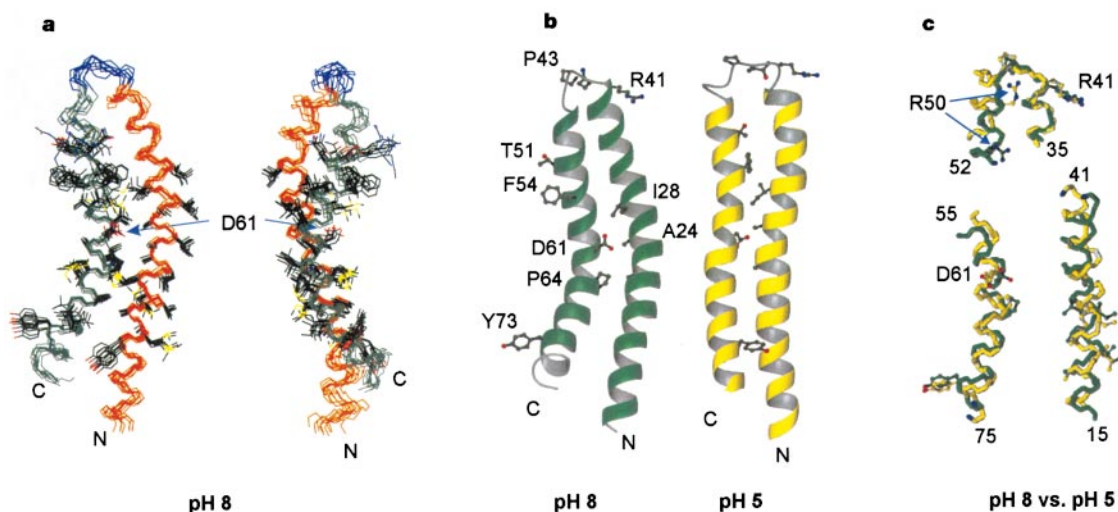


Figure 3 Structure of subunit *c* at pH 8 and comparison with the structure at pH 5. **a**, Two views of the backbone traces for the best nine structural models of subunit *c* at pH 8. The N-terminal helix is shown in orange, the polar loop in blue and the C-terminal helix in green. Side chains are shown for the transmembrane segments. **b**, Ribbon diagrams of deprotonated (green) and protonated (yellow) *c* subunits. Selected side chains are shown

to highlight the rotation of the C-terminal helix on deprotonation of Asp 61. **c**, Superpositions of the N-terminal helices (right), the polar loops (centre) and the C-terminal helices (left) of the Asp 61-deprotonated structure (green) and Asp 61-protonated structure (yellow). The figure was prepared using MOLMOL⁴⁹.

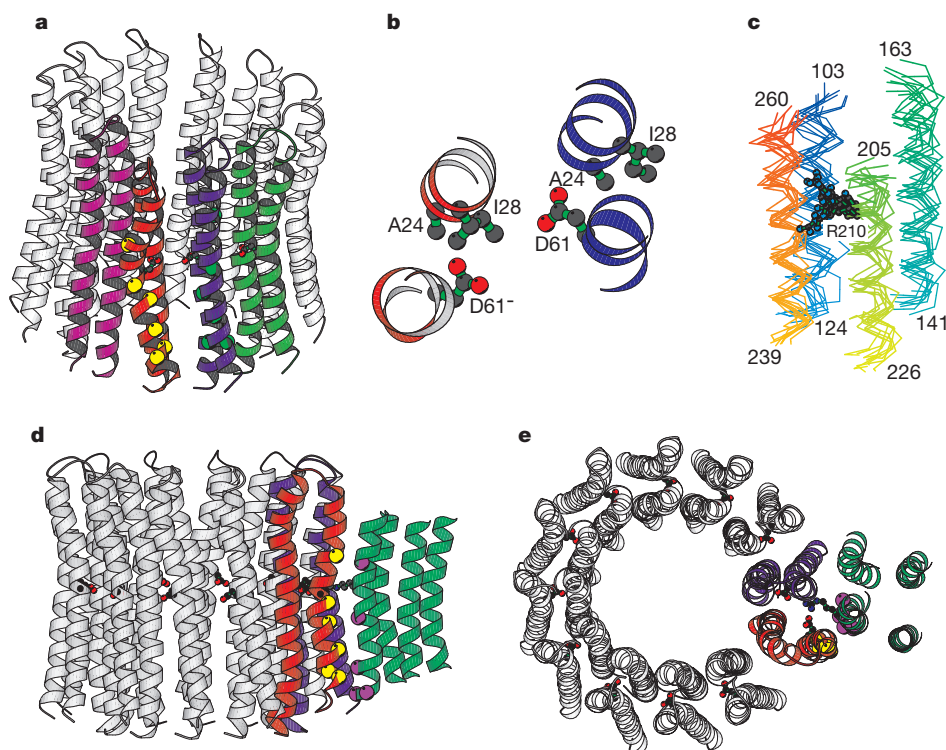


Figure 4 Models for the c_{12} and ac_{12} subcomplexes of F_0 . **a**, Model of the c_{12} oligomer, with the Asp 61-deprotonated subunit coloured in red, facing the viewer. The Asp 61 side chains are also shown for several subunits. Residues that can form disulphide cross links to subunit *a* are indicated with yellow spheres in the deprotonated subunit *c* and green spheres in the protonated subunit *c*. **b**, Region surrounding the deprotonated Asp 61 in the c_{12} complex. Ribbon traces show the backbone from residues 22–30 and 56–64. Note that the deprotonated Asp 61 side chain (on left) is rotated 140° clockwise from its

position in the adjacent monomer viewed from the polar loop. **c**, Alpha carbon traces of the best nine structural models of subunit *a*. Only the transmembrane segments and the Arg 210 side chains are shown. **d**, Side view of the ac_{12} complex. Subunit *a* is shown in green, the deprotonated subunit *c* in red and its protonated neighbour in purple. The residues that can form cysteine disulphide cross links are indicated as yellow spheres in subunit *c* and pink spheres in subunit *a*. **e**, Top view of ribbon diagram of the ac_{12} complex, with same colouring scheme as in **d**. The figure was prepared using Molscript⁵⁰.

204–226, and 238–263 of the four consensus helices; and helix–helix contacts between residues 119 → 245, 145 → 219, 210 → 252, 218 → 245 and 219 → 245, as summarized in ref. 39. An additional class of constraints was added to maintain the transmembrane helices in a roughly parallel orientation; the end of each helix was constrained to be at least as far from the opposite ends of adjacent helices as it was from the opposite end of itself. The resulting nine models that best satisfied the constraints are shown in Fig. 4c. This set of constraints defined the arrangement of the helices remarkably well, with r.m.s.d.s of 1.6 Å for the backbone atoms. When only the final two transmembrane helices were considered, for which the largest number of constraints were available, the r.m.s.d.s dropped to 1.1 Å. In these models the side chain of Arg 210 is positioned at the periphery of subunit *a*, where it could interact with the c_{12} oligomer. This method of calculating structural models from transmembrane helix topology and identified helical contacts should be of general use for the increasing number of membrane proteins for which such information is available. It could also form the basis for an iterative strategy: a set of initial constraints leading to a trial family of structural models could be used to predict possible contacts in the less ordered regions of the family, which would then be probed by pairwise cysteine substitutions, and the resulting new constraints used to calculate a better family of models.

To position subunit *a* with respect to the c_{12} oligomer, additional cysteine cross-linking data¹⁴ were used, which define the interactions between the penultimate helix of subunit *a* encompassing Arg 210 and the C-terminal helix of the deprotonated subunit *c* monomer of the c_{12} oligomer. Subunit *a* was positioned 75 Å from the c_{12} oligomer, and cysteine cross-linking constraints $a207 \rightarrow c55$, $a214 \rightarrow c62$, $a214 \rightarrow c65$, $a221 \rightarrow c69$, $a223 \rightarrow c72$ and $a224 \rightarrow c73$ from ref. 14 were used in a final set of torsion angle dynamics-

simulated annealing calculations to derive a working model for the ac_{12} complex (Fig. 4d, e). The resulting models satisfied all NMR and disulphide cross-linking constraints to within 0.4 Å, and were quite similar to each other with r.m.s.d.s between backbone atoms of 1.2 Å. The last two helices of subunit *a* packed against two subunit *c* monomers, with the Arg 210 side chain lying between the protonated and deprotonated Asp 61 side chains within the predicted H^+ translocation site.

Proton translocation and rotation within F_0

Proton translocation through F_0 involves residues Asp 61 of subunit *c* and Arg 210 of subunit *a* (Fig. 5a). As Asp 61 is buried in the membrane bilayer⁴¹, hydrophilic access pathways from both sides of the membrane must exist. One likely pathway connecting the periplasmic membrane surface to Arg 210 and Asp 61 can be found in the model of the ac_{12} oligomer. A set of polar residues entirely within subunit *a*, including Gln 252, Asn 214, Asn 148, Asp 119, His 245, Glu 219, Ser 144 and Asn 238, form a hydrophilic path between Arg 210 and the periplasmic membrane surface (Fig. 5b).

During ATP synthesis, the transmembrane H^+ electrochemical potential is high at the periplasmic face and low at the cytoplasmic (F_1) face. This gradient would drive protonation of the deprotonated Asp 61 via the pathway through subunit *a* just described. As this Asp 61 becomes protonated, its C-terminal helix will rotate, clockwise as viewed from F_1 in Fig. 5c, d, to adopt its stable protonated conformation. During this rotation, we suggest that subunit *a* moves with the C-terminal helix of *c*, linked by hydrogen bonding and steric interactions between residues in subunits *a* and *c*, to arrive at the next position between *c* subunits, as indicated in Fig. 5d, e. Although the structural models can only be treated as

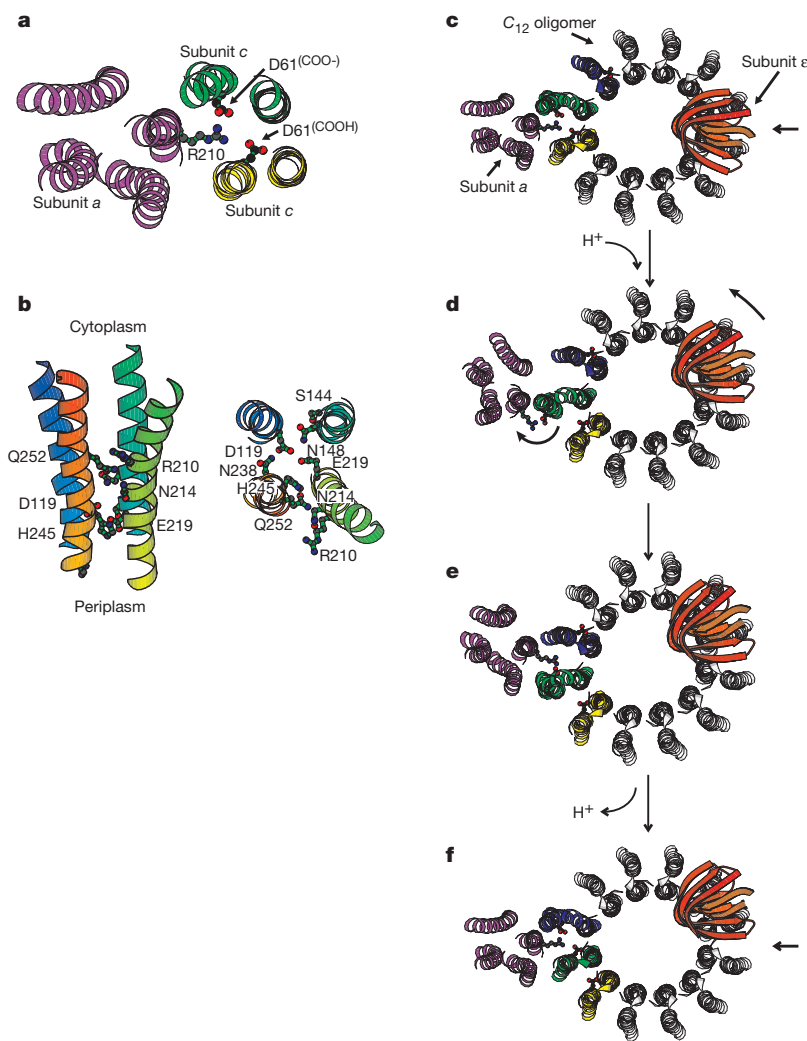


Figure 5 Proton translocation pathways and model for the functioning complex. **a**, The active site for proton translocation in the ac_{12} complex. **b**, Proposed proton path from the periplasmic membrane surface to Arg 210 of subunit *a* and Asp 61 of subunit *c*, viewed both from the subunit *c* interface (left) and down from the F_1 binding surface (right). **c–f**, Proposed functional cycle for translocation of one proton. **c**, Resting state. Arg 210 of subunit *a* lies between protonated and deprotonated Asp 61 side chains in the c_{12} oligomer. The location of subunit- ϵ is shown, positioned according to cross-linking data¹⁹, with its initial orientation indicated by an arrow at right. **d**, After protonation of Asp 61, the

C-terminal helix of the newly protonated monomer (shown in green) rotates towards its stable protonated orientation. **e**, Fully protonated intermediate. Subunit *a* is now at the interface to the next (purple) subunit *c*. The c_{12} ring has rotated by 30° with respect to subunit *a*. **f**, The Asp 61 of the next *c* subunit loses its proton to the F_1 side of the membrane, via a pathway involving both subunit *a* and subunit *c*. Its C-terminal helix rotates to adopt the stable conformation of the deprotonated state, regenerating the resting state of the enzyme. The initial position of subunit- ϵ is indicated by an arrow on the right, to highlight the rotation within the complex.

suggestive, potential hydrogen-bonding partners between the two subunits can be identified, including Asp 61 of *c* with Arg 210 of *a*, Arg 50 of *c* with Ser 202/206 of *a* and Thr 51 of *c* with Ser 202/206 of *a*. Bulky side chains in subunit *c* that would sterically impinge on subunit *a* include Phe 53, Met 57 and Met 65. The result of this concerted movement would be the rotation of the c_{12} oligomer by one-twelfth of a revolution with respect to subunit *a* (Fig. 5e). The Asp 61 of the next subunit *c* must then deprotonate to regenerate the starting configuration. Examination of the fully protonated intermediate state (Fig. 5e) suggests a potential hydrophilic path towards the F_1 side of the membrane. This access pathway is formed from polar residues in both subunit *a*, including Ser 206, Lys 203 and Ser 199, and two subunits *c*, including Arg 50 of one and Thr 51 and Gln 52 of the next. Deprotonating Asp 61 from the F_1 side of the membrane by this pathway would trigger rotation of the next C-terminal helix to its stable deprotonated conformation, regenerating the starting configuration (Fig. 5f).

This series of steps linked to translocation of a single proton leads to the rotation of the c_{12} oligomer by 30° with respect to the static elements a_1b_2 in F_0 and $\delta_1\alpha_3\beta_3$ in F_1 . The γ - and ϵ -subunits of the F_1

core are linked to the c_{12} ring through interactions with the polar loop of subunit c ^{18,19}, and would rotate in concert with the c_{12} ring^{24,42–44}. Four such steps would result in the observed 120° rotation of $\gamma\epsilon$ within the core of F_1 (refs 5–7, 45), which drives the binding changes required for catalysis^{3,4,24}.

This mechanism is the simplest one that follows directly from the conformational changes observed for subunit *c*, but it is not the only possibility. An alternative can be envisioned in which the c_{12} oligomer remains fixed with respect to subunit *a*, and a negative charge or 'proton hole' traverses the c_{12} ring, displacing the ϵ -subunit by 30°. In this mechanism, deprotonation of the first subunit *c* of the ring at the *a*–*c* interface would induce the rotation of its C-terminal helix, positioning the deprotonated Asp next to its protonated neighbour (counter-clockwise, as viewed from F_1). The proton would be exchanged between the two *c* subunits, passing a proton hole counter-clockwise. The process would repeat until it reached the *c* subunit underneath subunit- ϵ . There, the conformational change in the loop caused by helix rotation within subunit *c* would displace the ϵ -subunit from the loop of one subunit to the next as the proton hole passes, accomplishing the necessary rotation

of the $\epsilon\gamma$ -stalk. The proton exchanges would continue until reaching the last c subunit of the ring, next to subunit a . Finally, a proton would be taken up through the periplasmic pathway in subunit a (Fig. 5b), regenerating the starting state. Although this alternative runs counter to the generally predicted rotation of the c_{12} ring with respect to subunit a , it is more easily reconciled with some genetic data. In particular, the partial function of the Asp61Gly/Ala24Asp subunit c double mutant⁴⁶ is more easily accommodated by such a flickering helix mechanism, where either carboxylate-bearing helix could rotate to pass on the proton hole. However, until any structural information is available on the Asp 24/Gly 61 double mutant, its packing interactions and its conformational changes on deprotonation, we consider rotation of the c_{12} ring to be the most likely mechanism. In the double mutant, the C-terminal helix might still undergo rotation when Asp 24 deprotonates, and the other proposed hydrogen bonding and steric interactions between subunits c and a could suffice for the observed partial function.

Earlier models proposed simple rotational diffusion of a rigid c_{12} ring, possibly driven by electrostatic forces^{24,43,47}. The structural data on protonation-linked conformational changes in subunit c indicate that the process may be more mechanical, with local rotations within subunit c driving larger-scale rotations of the c_{12} oligomer as a whole, in a 'wheels within wheels' type of mechanism. The mechanism presented here was derived from the structural changes in subunit c observed by NMR, and an unbiased family of model structures for the ac_{12} complex derived from independent biochemical data. The actual proton access pathways will certainly be more intricate, and may even involve chains of water molecules as have been observed in crystal structures of several membrane proteins⁴⁸. The rotational steps and proton translocation pathways proposed here provide a good starting description of how proton translocation can be linked to rotation in F_1F_0 , and hence catalysis. Much more biochemical and structural work remains, however, to elucidate the details of this intriguing process. □

Received 26 July; accepted 21 September 1999.

1. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F_1 -ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
2. Bianchet, M. A., Hüllihen, J., Pedersen, P. L. & Amzel, L. M. The 2.8-Å structure of rat liver F_1 -ATPase: Configuration of a critical intermediate in ATP synthesis/hydrolysis. *Proc. Natl Acad. Sci. USA* **95**, 11065–11070 (1998).
3. Weber, J. & Senior, A. E. Catalytic mechanism of F_1 -ATPase. *Biochim. Biophys. Acta* **1319**, 19–58 (1997).
4. Boyer, P. The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
5. Duncan, T. M., Bulgin, V. V., Zhou, Y., Hutcheon, M. L. & Cross, R. L. Rotation of subunits during catalysis by *Escherichia coli* F_1F_0 ATPase. *Proc. Natl Acad. Sci. USA* **92**, 10964–10968 (1995).
6. Sabbert, D., Engelbrecht, S. & Junge, W. Intersubunit rotation in active F_1 -ATPase. *Nature* **381**, 623–625 (1996).
7. Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. J. Direct observation of the rotation of F_1 -ATPase. *Nature* **386**, 299–302 (1997).
8. Fillingame, R. H. In *The Bacteria*. Vol. XII. (ed. Krulwich, T. A.) 345–391 (Academic, New York, 1990).
9. Jones, P. C. & Fillingame, R. H. Genetic fusions of subunit c in the F_0 sector of H^+ -transporting ATP synthase. *J. Biol. Chem.* **273**, 29701–29705 (1998).
10. Birkenhäger, R., Hopper, M., Deckers-Hebestreit, G., Mayar, F. & Altendorf, K. The F_0 complex of the *Escherichia coli* ATP synthase: Investigation by electron spectroscopic imaging and immunoelectron microscopy. *Eur. J. Biochem.* **230**, 58–67 (1995).
11. Takeyasu, K. *et al.* Molecular imaging of *Escherichia coli* F_1F_0 -ATPase in reconstituted membranes using atomic force microscopy. *FEBS Lett.* **392**, 110–113 (1996).
12. Singh, S., Turina, P., Bustamante, C. J., Keller, D. J. & Capaldi, R. Topographical structure of membrane-bound *Escherichia coli* F_1F_0 ATP synthase in aqueous buffer. *FEBS Lett.* **397**, 30–34 (1996).
13. Jones, P. C., Jiang, W. & Fillingame, R. H. Arrangement of the multicopy H^+ -translocating subunit c in the membrane sector of the *Escherichia coli* F_1F_0 ATP synthase. *J. Biol. Chem.* **273**, 17178–17185 (1998).
14. Jiang, W. & Fillingame, R. H. Interacting helical faces of subunits a and c in the F_1F_0 ATP synthase of *Escherichia coli* defined by disulfide cross-linking. *Proc. Natl Acad. Sci. USA* **95**, 6607–6612 (1998).
15. Aris, J. P. & Simoni, R. D. Cross-linking and labeling of the *Escherichia coli* F_1F_0 -ATP synthase reveal a compact hydrophilic portion of F_0 close to an F_1 catalytic subunit. *J. Biol. Chem.* **258**, 14599–14609 (1983).
16. Hermolin, J., Gallant, J. & Fillingame, R. H. Topology, organization, and function of the psi subunit in the F_0 sector of the H^+ -ATPase of *Escherichia coli*. *J. Biol. Chem.* **258**, 14550–14555 (1983).
17. Rodgers, A. J. W. *et al.* The subunit δ -subunit b domain of the *Escherichia coli* F_1F_0 ATPase. *J. Biol. Chem.* **272**, 31058–31064 (1997).
18. Watts, S. D., Zhang, Y., Fillingame, R. H. & Capaldi, R. A. The γ subunit in the *Escherichia coli* ATP synthase complex (ECF₁F₀) extends through the stalk and contacts the c subunits of the F_0 part. *FEBS Lett.* **368**, 235–238 (1995).
19. Hermolin, J., Dmitriev, O. Y., Zhang, Y. & Fillingame, R. H. Defining the domain of binding of F_1 subunit ϵ with the polar loop of F_0 subunit c in the *Escherichia coli* ATP synthase. *J. Biol. Chem.* **274**, 17011–17016 (1999).

20. Fillingame, R. H., Girvin, M. E., Jiang, W., Valiyaveetil, F. & Hermolin, J. Subunit interactions coupling H^+ transport and ATP synthesis in F_1F_0 ATP synthase. *Acta Physiol. Scand.* **643**, 163–168 (1998).
21. Girvin, M. E., Rastogi, V. K., Abildgaard, F., Markley, J. L. & Fillingame, R. H. Solution structure of the transmembrane H^+ -transporting subunit c of the F_1F_0 ATP synthase. *Biochemistry* **37**, 8817–8824 (1998).
22. Matthey, U., Kaim, G., Braun, D., Wüthrich, K. & Dimroth, P. NMR studies of subunit c of the ATP synthase from *Propionigenium modestum* in dodecylsulfate micelles. *Eur. J. Biochem.* **261**, 459–467 (1999).
23. Norwood, T. J., Crawford, D. A., Steventon, M. E., Driscoll, P. C. & Campbell, I. D. Heteronuclear 1H - ^{15}N nuclear magnetic resonance studies of the c subunit of the *Escherichia coli* F_1F_0 ATP synthase: assignment and secondary structure. *Biochemistry* **31**, 6285–6290 (1992).
24. Junge, W., Lill, H. & Engelbrecht, S. ATP synthase: an electrochemical transducer with rotatory mechanics. *Trends Biochem. Sci.* **22**, 420–423 (1997).
25. Zuideweg, E. R. P. & Fesik, S. W. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28**, 2387–2391 (1989).
26. Assadi-Porter, F. M. & Fillingame, R. H. Proton-translocating carboxyl of subunit c of $F_1F_0H^+$ -ATP synthase: the unique environment suggested by the pK_a determined by 1H NMR. *Biochemistry* **34**, 16186–16193 (1995).
27. Rastogi, V. K. & Girvin, M. E. 1H , ^{13}C , and ^{15}N assignments and secondary structure of the high pH form of subunit c of the F_1F_0 ATP synthase. *J. Biomol. NMR* **13**, 91–92 (1999).
28. Stein, E. G., Rice, L. M. & Brünger, A. T. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.* **124**, 154–164 (1997).
29. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
30. Baxter, N. J. & Williamson, M. P. Temperature dependence of 1H chemical shifts in proteins. *J. Biomol. NMR* **9**, 359–369 (1997).
31. Kuszewski, J., Qin, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **106**, 92–96 (1995).
32. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **107**, 293–297 (1995).
33. Dmitriev, O. Y., Jones, P. C. & Fillingame, R. H. Modeling the oligomeric arrangement of subunit c in the F_1F_0 ATP synthase from the structure of the monomeric subunit and cross-linking in the native enzyme. *Proc. Natl Acad. Sci. USA* **96**, 7785–7790 (1999).
34. Groth, G. & Walker, J. E. Model of the c -subunit oligomer in the membrane domain of F_1 -ATPases. *FEBS Lett.* **410**, 117–123 (1997).
35. Girvin, M. E. & Fillingame, R. H. Determination of local protein structure by spin label difference 2D NMR: the region neighboring Asp61 of subunit c of the F_1F_0 ATP synthase. *Biochemistry* **34**, 1635–1645 (1995).
36. Valiyaveetil, F. I. & Fillingame, R. H. On the role of Arg-210 and Glu-219 of subunit a in proton translocation by the *Escherichia coli* F_1F_0 -ATP synthase. *J. Biol. Chem.* **272**, 32635–32641 (1997).
37. Yamada, H., Moriyama, Y., Maeda, M. & Futai, M. Transmembrane topology of *Escherichia coli* H^+ -ATPase (ATP synthase) subunit a . *FEBS Lett.* **290**, 34–38 (1996).
38. Jager, H., Birkenhäger, R., Stalz, W. D., Altendorf, K. & Deckers-Hebestreit, G. Topology of subunit a of the *Escherichia coli* ATP synthase. *Eur. J. Biochem.* **251**, 122–132 (1998).
39. Valiyaveetil, F. I., Fillingame, R. H. Transmembrane topography of subunit a in the *Escherichia coli* F_1F_0 ATP synthase. *J. Biol. Chem.* **273**, 16241–16247 (1998).
40. Wada, T., Long, J. C., Zhang, D. & Vik, S. B. A novel labeling approach supports the five-transmembrane model of subunit a of the *Escherichia coli* ATP synthase. *J. Biol. Chem.* **274**, 17353–17357 (1999).
41. Sigrist-Nelson, K. & Azzi, A. The proteolipid subunit of chloroplast adenosine triphosphatase complexes: mobility, accessibility, and interactions by a spin label technique. *J. Biol. Chem.* **254**, 4470–4474 (1979).
42. Vik, S. B. & Antonio, B. J. A mechanism of proton translocation by F_1F_0 ATP synthases suggested by double mutants of the a subunit. *J. Biol. Chem.* **269**, 30364–30369 (1994).
43. Elston, T., Wang, H. & Oster, G. Energy transduction in ATP synthase. *Nature* **391**, 510–513 (1998).
44. Howitt, S. M., Rodgers, A. J. W., Hatch, L. P., Gibson, F. & Cox, G. B. The coupling of the relative movement of the a and c subunits of the F_0 to the conformational changes in the F_1 -ATPase. *J. Bioenerg. Biomemb.* **28**, 415–420 (1996).
45. Kato-Yamada, Y., Noji, H., Yasuda, R., Kinosita, K. & Yoshida, M. Direct observation of the rotation of the ϵ subunit in F_1 -ATPase. *J. Biol. Chem.* **273**, 19375–19377 (1998).
46. Miller, M. J., Oldenburg, M. & Fillingame, R. H. The essential carboxyl group in subunit c of the F_1F_0 ATP synthase can be moved and H^+ translocating function retained. *Proc. Natl Acad. Sci. USA* **87**, 4900–4904 (1990).
47. Dimroth, P., Wang, H., Grabe, M. & Oster, G. Energy transduction in the sodium F_1 -ATPase of *Propionigenium modestum*. *Proc. Natl Acad. Sci. USA* **96**, 4735–4737 (1999).
48. Ermler, U., Fritzsche, G., Buchanan, S. K. & Michel, H. Structure of the photosynthetic reaction centre from *Rhodospirillum rubrum* at 2.65 Å resolution: cofactors and protein-cofactor interactions. *Structure* **2**, 925–936 (1994).
49. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for the display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55 (1996).
50. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

Acknowledgements

We thank S. Cahill for assistance with the NMR experiments, and M. Brenowitz and V. Schramm for critical reading and discussion of the manuscript.

Correspondence and requests for materials should be addressed to M.E.G. (e-mail: girvin@aecom.yu.edu). Atomic coordinates are available at the Protein Data Bank for the structure of subunit c at pH 8 (1c99), the refined structure of subunit c at pH 5 (1c0v), and the model for the ac_{12} complex (1c17). Chemical shifts are available at the BioMagResBank for subunit c at pH 8 (entry 4151) and pH 5 (entry 4146).

A low-temperature origin for the planetesimals that formed Jupiter

Tobias Owen*, Paul Mahaffy†, H. B. Niemann†, Sushil Atreya‡, Thomas Donahue‡, Akiva Bar-Nun§ & Imke de Pater||

* University of Hawaii, Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

† Laboratory for Atmospheres, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

‡ Atmospheric Ocean & Space Science, University of Michigan, Ann Arbor, Michigan 48109, USA

§ Department of Geophysics & Planetary Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel

|| Department of Astronomy, University of California, Berkeley, 601 Campbell Hall, Berkeley, California 94720-3411, USA

The four giant planets in the Solar System have abundances of 'metals' (elements heavier than helium), relative to hydrogen, that are much higher than observed in the Sun. In order to explain this, all models for the formation of these planets rely on an influx of solid planetesimals¹⁷. It is generally assumed that these planetesimals were similar, if not identical, to the comets from the Oort cloud that we see today. Comets that formed in the region of the giant planets should not have contained much neon, argon and nitrogen, because the temperatures were too high for these volatile gases to be trapped effectively in ice. This means that the abundances of those elements on the giant planets should be approximately solar. Here we show that argon, krypton and xenon in Jupiter's atmosphere are enriched to the same extent as the other heavy elements, which suggests that the planetesimals carrying these elements must have formed at temperatures lower than predicted by present models of giant-planet formation.

We are continuing the calibration of the Galileo probe mass spectrometer using a duplicate of the flight model instrument. We previously reported revised values of ³He/⁴He and D/H from this work¹. Proceeding to the noble gases², we have now found a value of 2.1 ± 0.4 times the solar value for the argon mixing ratio, clearly indicating enrichment over the expected solar nebular value. We have also found krypton and xenon to be similarly enhanced, at 2.7 ± 0.5 and 2.6 ± 0.5 times the solar values, respectively. These new results supercede our previously published upper limits (≤ 5 times the solar value) for ⁸⁴Kr and ¹³²Xe, and our preliminary evaluation of the argon mixing ratio at 2.5 ± 0.5 times the solar value³. Aside from the neon that dissolves in the 'raindrops' of helium that are coming out of solution in the fluid metallic hydrogen in the planet's interior⁴, it appears that the heavy noble gases share essentially the same enrichment as carbon and sulphur on Jupiter (Fig. 1).

This concordance invites a review of the jovian nitrogen abundance, as ground-based microwave observations have suggested that N/H is only 1.2–1.3 times the solar value^{5–8}. Although we have not yet been able to obtain more than an upper limit on nitrogen from the mass spectrometer, Folkner *et al.*⁹ succeeded in deriving N/H from analysis of the attenuation of the probe's radio signal by NH₃. Adjusting this determination⁹ of the jovian ammonia mixing ratio to the solar abundances established by Anders and Grevesse¹⁰ (which we use throughout), we find $N/H = 3.6 \pm 0.5$ times the solar value at pressure levels below approximately 8–10 bar. This is consistent with our new results for noble gases as illustrated in Fig. 1. We believe the apparent discrepancy between the probe and ground-based ammonia mixing ratios can be resolved by examining the pressure levels sampled by the microwave observations and the error bars associated with these measurements. We have showed¹¹ that it is possible to reconcile the Folkner *et al.*⁹ profile of ammonia

with the ground-based microwave spectrum of the planet, provided ammonia is depleted at pressure levels above ~ 4 bar.

The cause of the ammonia depletion remains unknown. It is not the result of a 10 times solar abundance of sulphur forming thick NH₄SH clouds, as de Pater⁶ originally suggested, because sulphur, like carbon, is only enriched by about a factor of 3 in Jupiter's atmosphere. Some type of meteorological process seems more likely. If H₂S and H₂O both behave as NH₃ does in the globally averaged jovian atmosphere, these species will only reach their maximum mixing ratios at altitudes considerably below the equilibrium condensation levels corresponding to ~ 3 times solar abundances.

We conclude that the deep, well-mixed atmosphere of Jupiter exhibits essentially the same enrichment of nitrogen as it does for carbon, sulphur, the heavy noble gases, and, we assume, everything else except neon and helium. This requires that argon, krypton, xenon and nitrogen were present in solar proportions relative to carbon and sulphur in the icy planetesimals that contributed to Jupiter's formation. We know of no solid materials in the Solar System that exhibit this composition, although we have suggested it for the Kuiper belt objects¹².

This suggestion was based on laboratory studies of the trapping of highly volatile gases in amorphous ice forming at temperatures below 75 K (refs 12, 13). These experiments show that trapping argon and N₂ in solar abundances relative to carbon in icy planetesimals requires condensation of the ice at temperatures ≤ 30 K (capturing neon in this proportion requires a temperature $T < 17$ K). Such temperatures would be appropriate to the region of the original solar nebula beyond the orbit of Pluto, where we now find the Kuiper belt. In contrast, the Oort cloud comets are thought to have formed in the Uranus–Neptune region, where solar nebula temperatures were of the order of 55 K (ref. 14). Ice forming there could only have trapped 1% of the ambient argon and N₂. It is generally accepted that $\geq 70\%$ of interstellar (and hence outer solar

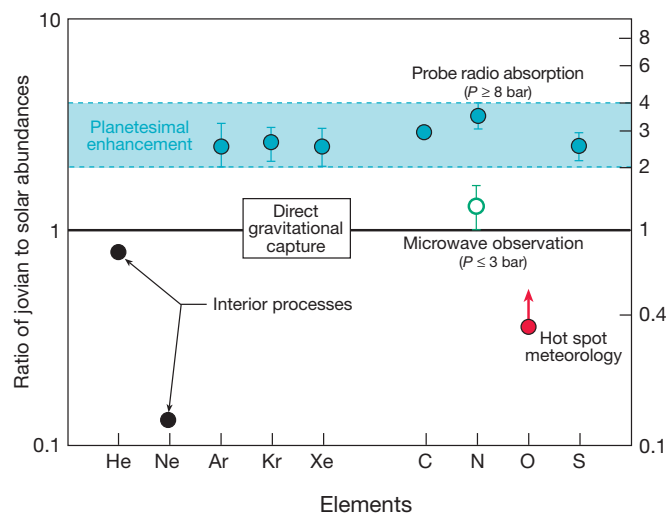


Figure 1 Elemental abundances (relative to hydrogen) in Jupiter's atmosphere compared with solar values. All the elements except nitrogen were measured by the Galileo probe mass spectrometer. The nitrogen abundance was determined by Folkner *et al.*⁹ from the attenuation of the probe radio signal by ammonia in Jupiter's atmosphere. These data were derived from calibrations carried out with a duplicate model of the mass spectrometer used on the Galileo probe, known as the Engineering Unit. These calibrations may be more accurate, in many cases, than those carried out before launch on the Flight Unit, as they exercise the full range of the instrument's capabilities—including the enrichment cells—under conditions that duplicate those encountered during the descent. For example, gas mixtures are used in the Engineering Unit calibrations that are more representative of the actual jovian atmosphere than mixtures used to calibrate the Flight Unit before launch and the temperatures realized by enrichment cells during the descent can be duplicated in these experiments. Recent refurbishment has restored the Engineering Unit to a condition that even more closely resembles that of the Flight Unit.

nebula) nitrogen is in the form of N_2 (refs 15–17); this explains the observed deficiency of nitrogen (argon has not yet been detected) in Oort cloud comets^{18–20}.

It seems to us that the only explanation for the uniform enrichment of heavy elements that we observe on Jupiter is that these elements came to the planet in very cold ($T < 30$ K) icy planetesimals. This differs from conventional models, which relate the formation of giant planets to the threshold distance from their stars where it first becomes cold enough for water ice to condense in the circumstellar disks. In these 'snowline' models, the ice condenses at $T \approx 160$ K (~ 5 AU in our Solar System)^{14,17}. Argon and N_2 would be depleted by over 4 orders of magnitude in icy planetesimals formed at such temperatures¹².

We need to devise ways in which such low temperature material could have been produced and transported to the giant planets. Could planetesimals have formed in the interstellar medium during the early phases of cloud fragmentation and collapse? Or was the solar nebula much colder than current models predict? Or did Jupiter migrate to its present position from beyond 30 AU? The difficulties inherent in each of these suggestions emphasize the significance of this new constraint.

Additional tests will require more probes to the outer planets. We predict that the global oxygen abundance on Jupiter should show the same enrichment as the other elements. The apparent depletion in Fig. 1 must result from the same local, meteorological effect acting on H_2O , that has depleted H_2S and NH_3 in the upper part of the probe's trajectory^{21–23}. Perhaps all the outer planets show the same enhancement of Ar and N as we find on Jupiter. The presence of HCN in the upper atmosphere of Neptune may indicate that this is the case, as this gas appears to require planetary N_2 for its existence²⁴. □

Received 26 April; accepted 1 October 1999.

- Mahaffy, P. R., Donahue, T. M., Atreya, S. K., Owen, T. & Niemann, H. B. in *Primordial Nuclei and their Galactic Evolution* (eds Prantzos, N., Tosi, M. & von Steiger, R.) 239–250 (Kluwer, Dordrecht, 1998).
- Mahaffy, P. R., Niemann, H. B., Alpert, A., Atreya, S. K., Donahue, T. M. & Owen, T. Heavy noble gases in the atmosphere of Jupiter. *Bull. Am. Astron. Soc.* **30**, 1066 (1998).
- Niemann, H. *et al.* The composition of the Jovian atmosphere as determined by the Galileo probe mass spectrometer. *J. Geophys. Res.* **103**, 22831–22845 (1998).
- Roulston, M. S. & Stevenson, D. J. Prediction of neon depletion in Jupiter's atmosphere (abstract). *Eos* **76**, F343 (1995).
- de Pater, I. & Massie, S. T. Models of the millimeter-centimeter spectra of the giant planets. *Icarus* **62**, 143–171 (1985).
- de Pater, I. Jupiter's zone-belt structure at radio wavelengths. *Icarus* **68**, 344–365 (1986).
- Joiner, J. & Steffes, P. B. Modelling of Jupiter's millimeter wave emission utilizing laboratory measurements of ammonia (NH_3) opacity. *J. Geophys. Res.* **96**, 17463–17470 (1991).
- Romani, P. N., de Pater, I. & Zahnle, K. Galileo, SL9 and Jupiter's deep atmosphere. *Bull. Am. Astron. Soc.* **27**, 81 (1995).
- Folkner, W. M., Woo, R. & Nandi, S. Ammonia abundance in Jupiter's atmosphere derived from attenuation of the Galileo probe's radio signal. *J. Geophys. Res.* **103**, 22847–22856 (1998).
- Anders, E. & Grevesse, N. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Atreya, S. K., Wong, M. H., Owen, T., Mahaffy, P. R., Niemann, H., de Pater, I., Drossart, P. & Encreanz, Th. A comparison of the atmosphere of Jupiter and Saturn: Deep atmospheric composition, cloud structure, vertical mixing and origin. *Planet. Space Sci.* (in the press).
- Owen, T. & Bar-Nun, A. Comets, impacts and atmospheres. *Icarus* **116**, 215–226 (1995).
- Bar-Nun, A., Kleinfeld, I. & Kochavi, E. Trapping of gas mixtures by amorphous water ice. *Phys. Rev. B* **38**, 7749–7754 (1988).
- Cassen, P. Utilitarian models of the solar nebula. *Icarus* **112**, 405–429 (1994).
- Linke, R. A., Guelin, M. & Langer, W. O. Detection of $H^{15}NN^+$ and $HN^{15}N^+$ in interstellar clouds. *Astrophys. J.* **371**, L85–L88 (1983).
- Womack, M., Wyckhoff, S. & Ziurys, L. M. Observational constraints on Solar Nebula nitrogen chemistry: N_2/NH_3 . *Astrophys. J.* **401**, 728–735 (1992).
- Van Dishoeck, E. F., Blake, G. A., Draine, B. T. & Lunine, J. I. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 163–244 (Univ. Arizona Press, Tucson, 1993).
- Geiss, J. Composition of Halley's Comet: Clues to origin and history of cometary matter. *Rev. Modern Astron.* **1**, 1–27 (1988).
- Krankowsky, D. in *Comets in the Post-Halley Era* (eds Newburn, R. L. Jr, Neugebauer, M. & Rahe, J.) 855–878 (Kluwer, Dordrecht, 1991).
- Wyckoff, S., Tegler, S. C. & Engel, L. Nitrogen abundance in Comet Halley. *Astrophys. J.* **367**, 641–648 (1991).
- Owen, T., Atreya, S. K., Mahaffy, P., Niemann, H. B. & Wong, M. H. in *Three Galileos: The Man, The Spacecraft, The Telescope* (eds Barbieri, C., Rahe, J., Johnson, T. V. & Sohus, A.) 289–297 (Kluwer, Dordrecht, 1997).
- Atreya, S. K., Owen, T. C., Wong, M. H., Niemann, H. B. & Mahaffy, P. R. in *Three Galileos: The Man, The Spacecraft, The Telescope* (eds Barbieri, C., Rahe, J., Johnson, T. V. & Sohus, A.) 289–297 (Kluwer, Dordrecht, 1997).

- Showman, A. P. & Ingersoll, A. P. Interpretation of Galileo Probe data and implications for Jupiter's dry downdrafts. *Icarus* **132**, 205–220 (1998).
- Marten, A. *et al.* First observations of CO and HCN on Neptune and Uranus at millimeter wavelengths and their implications for atmospheric chemistry. *Astrophys. J.* **406**, 285–297 (1993).

Acknowledgements

We thank NASA for supporting this research.

Correspondence and requests for materials should be addressed to T.O.
(e-mail: owen@ifa.hawaii.edu).

Lithium nucleosynthesis in the Sun inferred from the solar-wind $^7Li/^6Li$ ratio

Marc Chaussidon* & François Robert†

* CRPG-CNRS, BP 20, 54501 Vandoeuvre-lès-Nancy Cedex, France

† MNHN-CNRS, 61 rue Buffon, 75005 Paris, France

The abundance of lithium measured in meteorites has generally been assumed to be the 'Solar System value', which presumably reflects the abundance in the gas cloud out of which the Sun formed¹. Lithium is a factor of 140 less abundant in the solar photosphere than in meteorites²; this difference has been attributed to the destruction of lithium (through nuclear reactions) at the base of the Sun's convection zone^{3–5}. If this is correct, then the ratio of $^7Li/^6Li$ in the Sun's photosphere should be $\sim 10^6$ (ref. 6), as 6Li is destroyed much more easily (at a lower temperature) than 7Li : the meteoritic abundance ratio is $^7Li/^6Li = 12.14$ (ref. 7). Here we report that $^7Li/^6Li = 31 \pm 4$ for lithium in the solar wind that has been implanted in lunar soil. This low ratio suggests that lithium is produced when energetic protons from solar flares induce spallation reactions with ^{16}O and ^{12}C present in the photosphere.

Solar-wind particles, because of their low energy (~ 8 keV), are implanted within the first $\sim 0.03 \mu m$ of the grains constituting the lunar soils⁸. With the exception of volatile elements^{9,10}, the isotopic composition of the solar wind has never been measured because of the difficulty of extracting selectively the solar fraction of non-volatile elements from lunar rocks. We describe here an approach to this technical problem, using ion-probe shallow-depth profiling for the determination of the $^7Li/^6Li$ ratio of the trapped solar wind. Four samples of lunar soils (79261, 79035, 79221, 10060) have been analysed for Li (19 depth profiles corresponding to $\sim 4,000$ determinations of the $^7Li/^6Li$ ratios and Li concentrations). Examples of depth profiles are shown in Fig. 1. The total depth of each profile, and the maximum and minimum measured $^7Li/^6Li$ ratios, are reported in Table 1.

In the soil samples marked departures of the $^7Li/^6Li$ ratio from the Solar System range⁷ are observed as a function of the sputtering time. With one exception (sample 10060,4), Li sputtered from the first $0.01 \mu m$ is enriched in 7Li ($^7Li/^6Li$ ratios up to 22.8) compared to that from further in. Since the solar wind Li (SW Li) should be enriched in 7Li due to the rapid astration (Li destruction in stars) of 6Li compared to 7Li (refs 6, 11), this surface 7Li -rich component can be ascribed to solar-wind implantation. The low $^7Li/^6Li$ ratios (down to 6.8) are generally observed deeper than SW Li. This 6Li -rich component can be ascribed to spallation reactions (spall-Li) taking place through the interaction between lunar soils and the solar (or galactic) energetic cosmic rays which yields¹² a $^7Li/^6Li$ ratio close to 2. At greater depth the measured $^7Li/^6Li$ ratios approach 12,

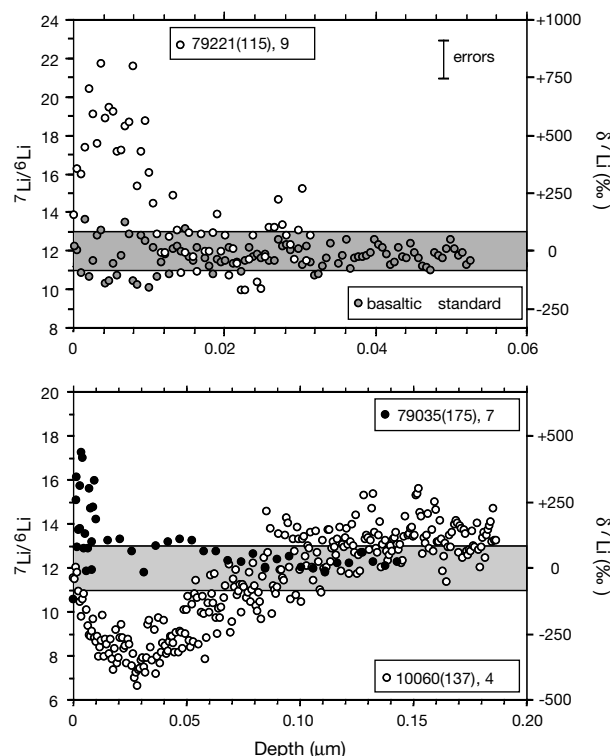


Figure 1 The $^7\text{Li}/^6\text{Li}$ isotopic ratio as a function of depth for three lunar soils (labelled by sample number). The Li isotopic ratios are also given in $\delta^7\text{Li}$ units, that is, in ‰ variations relative to the Earth's mantle. The terrestrial basaltic standard is shown for comparison: its $^7\text{Li}/^6\text{Li}$ ratio (12.14) is constant within $\pm 8\%$, corresponding to the counting statistics on ^6Li (8 s counting at ~ 20 counts s^{-1}). A marked increase in $^7\text{Li}/^6\text{Li}$ is found in the first $0.01 \mu\text{m}$ of the soils: this is attributed to solar-wind implantation. The component with a low $^7\text{Li}/^6\text{Li}$ ratio, apparent at $\sim 0.04 \mu\text{m}$ depth in soil 10060, is attributed to *in situ* spallation. Depth profiles were obtained on soil samples pressed in indium with an ims 3f ion microprobe using an O^- primary beam. The samples were sputtered over

$250 \times 250 \mu\text{m}^2$ ($\sim 10^3$ – 10^4 grains) and the dynamic transfer (for refocalization of secondary ions) was used. The average sputtered depth is $5.3 \times 10^{-4} \mu\text{m}$ between each data point (18 s). For this sputtering rate ($4.2 \times 10^{-3} \text{ nm nA}^{-1} \text{ s}^{-1}$), 10^2 – 10^3 atoms of ^6Li are sputtered each 18 s for a Li concentration of 1 p.p.b. Most of the data reported here were collected for ^6Li intensities ≥ 10 counts s^{-1} , thus corresponding to precisions $\leq \pm 10\%$ on the $^7\text{Li}/^6\text{Li}$ ratios (corrected for a 41‰ mass discrimination). The calibration of the variations of the Si^+/Li^+ ratio versus Si^+ intensity during depth profiling of standard glasses allows the determination on the samples, at each depth, of the absolute Si/Li ratio within $\pm 25\%$. Li concentrations are expressed in p.p.m.

which is the chondritic and terrestrial ratio. This lithium component is derived from the lunar mantle (lunar Li). Thus, the $^7\text{Li}/^6\text{Li}$ ratio of the pure solar wind can be estimated by a mixing model involving three end members: SW Li, lunar Li and spall-L Li (Fig. 2). The four samples reported in Table 1 give a mean SW $^7\text{Li}/^6\text{Li}$ ratio of 31 ± 4 . We note that this ratio corresponds to the spectroscopic lower-limit ratio ($^7\text{Li}/^6\text{Li} > 33$) measured in sunspots¹³.

The concentrations of SW Li and spall-L Li can also be calculated from the profiles. The following criteria were used for this calculation: (1) Li with $11.0 < ^7\text{Li}/^6\text{Li} < 13.0$ was ascribed to lunar Li, (2) the concentration of SW Li was calculated for $^7\text{Li}/^6\text{Li} > 13.0$ by a mass balance between lunar Li ($^7\text{Li}/^6\text{Li} = 12.14$) and SW Li ($^7\text{Li}/^6\text{Li} = 31$), (3) the concentration of spall-L Li was calculated for $^7\text{Li}/^6\text{Li} < 11.0$ by a mass balance between lunar Li and spall-L Li ($^7\text{Li}/^6\text{Li} = 2.0$). According to these calculations, the surface concentration within two atomic layers of lunar grains can exceed 1 p.p.m. for SW Li and is close to 1 p.p.m. for spall-L Li. At this scale, the two components appear intimately mixed, as expected from the complexity of the thin coatings which cover lunar soil grains and their complex maturation history^{14–16}. This distribution is also a clear indication that the spall-L Li has been mobilized from its deep ($\sim 1 \text{ cm}$) production site in the lunar basalt¹⁷. Supporting this interpretation, the calculated exposure ages seem meaningless (see the parameters of calculations in Table 1).

The present determination of a SW $^7\text{Li}/^6\text{Li}$ ratio of 31 ± 4 contradicts the standard models of the solar photosphere. Indeed, to account for the lithium depletion in the Sun (present-day photospheric $\text{Li}/\text{H} = 1.45 \times 10^{-11}$) compared to the proto-solar nebula (chondritic $\text{Li}/\text{H} = 210 \times 10^{-11}$), it is generally admitted

that Li is converted to He at the base of the convection zone (that is, at a depth of 0.28 times the solar radius¹⁸ where the temperature is close to $2.5 \times 10^6 \text{ K}$), and that the remaining Li is almost free of ^6Li (refs 6, 11, 19; $^7\text{Li}/^6\text{Li} = 10^6$). This discrepancy between the theoretical and the observed solar $^7\text{Li}/^6\text{Li}$ ratios has three possible implications for solar physics.

First, that the nuclear reaction rates for Li destruction are incorrect. Assuming a first-order rate for the destruction of Li in the Sun (that is, $d[^m\text{Li}]/dt = -k_m[\text{Li}]$, with m standing for mass 6 or 7), the evolution of the Li/H ratio from the proto-solar value to the present-day photospheric value can be calculated. Accordingly, ^6Li should be destroyed in the Sun only 1.2 times more rapidly than ^7Li in order to increase the $^7\text{Li}/^6\text{Li}$ ratio from 12.14 to 31.0. Such an estimate seems highly unreasonable, because it would imply that the theoretical destruction rate for ^6Li relative to ^7Li is too high by two orders of magnitude¹¹.

Second, that solar convection is more complex than previously anticipated. The present results would require that most of the left-over photospheric Li has never been buried at depth in the Sun. In addition, in order to maintain a low photospheric $^7\text{Li}/^6\text{Li}$ ratio ($\leq 31 \pm 4$), no ^7Li should be added to the photosphere via upwards convection currents. This requires a surficial convection decoupled from the lower one. However, this interpretation seems difficult to reconcile with the solar Be/H ratio (ref. 20) and the solar-wind $^{11}\text{B}/^{10}\text{B}$ ratios (measured along with the $^7\text{Li}/^6\text{Li}$ ratio) which are both chondritic²¹ ($3 < ^{11}\text{B}/^{10}\text{B} < 5$; two depth profiles not shown here). These observations would imply a narrow range for the depth of the convection zone (that is, a restricted range for the destruction ratio of ^6Li relative to ^7Li) where ^6Li and ^7Li are totally destroyed while no

Table 1 Concentrations and isotopic compositions of Li in lunar soils

Lunar sample*	Depth† (μm)	Solar wind Li (SW Li)‡			Lunar spallogenic Li (Spall-L Li)§		Age (Gyr)
		[SW Li] (p.p.m.)	$^7\text{Li}/^6\text{Li}_{\text{max}}$	$^7\text{Li}/^6\text{Li}_{\text{SW}}$	[Spall-L Li] (p.p.m.)	$^7\text{Li}/^6\text{Li}_{\text{min}}$	
79221(115), 4	0.030	0.64	15.1	ND	<0.001	11.7	–
79221(115), 6	0.053	0.10	17.5	ND	0.003	10.8	0.3
79221(115), 7	0.057	0.023	16.0	ND	0.029	7.3	3.2
79221(115), 9	0.031	1.74	21.7	29 ± 3	0.005	10.2	0.5
79035(175), 1	0.037	0.088	16.3	ND	0.015	9.1	1.7
79035(175), 4	0.079	0.15	18.0	34 ± 5	0.001	9.1	0.2
79035(175), 6	0.021	0.16	22.8	33 ± 3	0.001	10.7	0.2
79035(175), 7	0.010	0.42	17.1	ND	0.001	10.7	0.2
10060(137), 3	0.040	0.050	14.6	ND	0.032	9.6	3.5
10060(137), 4	0.187	0.14	15.5	26 ± 3	0.108	6.8	(12.0)
10060(137), 5	0.111	0.032	13.4	ND	0.001	10.8	0.2
79261(97), 1	0.107	<0.001	12.4	ND	0.009	10.8	1.0
79261(97), 2	0.106	0.033	13.5	ND	0.087	9.2	(9.7)
79261(97), 5	0.096	<0.001	12.4	ND	0.007	10.7	0.8
79261(97), 8	0.096	0.008	13.7	ND	0.012	10.9	1.3
79261(97), 9	0.096	0.007	13.7	ND	0.019	10.8	2.2
79261(97), 11	0.128	<0.001	13.3	ND	0.020	10.4	2.2
79261(97), 12	0.123	<0.001	12.5	ND	0.022	10.7	4.2
79261(97), 14	0.061	0.028	13.4	ND	<0.001	10.6	–

These data on solar-wind Li and lunar spallogenic Li were obtained with an ion microprobe. ND, not determined as the error propagation on SW $^7\text{Li}/^6\text{Li}$ exceeded $\pm 50\%$.

* Last numbers refer to different depth profiles in the same lunar sample.

† Depth of the $250 \times 250 \mu\text{m}^2$ ion probe craters. The data for a given sample are obtained by the summation of individual determinations performed each 18 s corresponding to an average depth of $5.3 \times 10^{-4} \mu\text{m}$ (see depth profiles in Fig. 1).

‡ Values of $^7\text{Li}/^6\text{Li} > 13.0$ were attributed to the presence of a detectable solar-wind contribution whose isotopic composition is taken to be equal to 31. Its concentration is calculated in the volume defined by the sputtering depth for a rock density = 1.5 g cm^{-3} . The maximum measured $^7\text{Li}/^6\text{Li}$ in each depth profile is reported in column 4. The $^7\text{Li}/^6\text{Li}$ of the solar wind (column 5) is obtained from a three-component mixing model between lunar Li ($^7\text{Li}/^6\text{Li} = 12.1$), spall-L Li ($^7\text{Li}/^6\text{Li} = 2.0$) and SW Li (see Fig. 2).

§ Values of $^7\text{Li}/^6\text{Li} < 11.0$ were attributed to the presence of a detectable spallogenic lunar contribution whose isotopic composition is equal to 2.0. Its concentration is calculated similarly to that of SW Li. The corresponding exposure ages are calculated using a solar flux of $70 \text{ H cm}^{-2} \text{ s}^{-1}$ ($E > 10 \text{ MeV}$), 15 mbarn for $^{16}\text{O}(p,x)^7\text{Li}$ and 41 wt% ^{16}O in the lunar basalt.

destruction takes place for Be and B (burning at $T \geq 3 \times 10^6 \text{ K}$; ref. 4).

Third, that there is a nuclear source of Li in the Sun. Spallation reactions occurring in the Sun's atmosphere, resulting from interactions between solar flares and ^{16}O (and ^{12}C), can in principle produce Li ($^7\text{Li}/^6\text{Li} = 2$). Such a process has previously been proposed to account for the $^7\text{Li}/^6\text{Li}$ ratio measured in low-metallicity stars (for example, $^7\text{Li}/^6\text{Li} = 19 \pm 8$; ref. 22). Spallation reactions in the Sun have also been invoked to explain the excess of solar-wind ^{10}Be and ^{14}C implanted in lunar soils^{23,24}. The fraction of spallogenic Li in the Sun's atmosphere can be bracketed using these proposed $^{10}\text{Be}/\text{H}$ (1×10^{-14}) and $^{14}\text{C}/\text{H}$ ($(2.5\text{--}3.5) \times 10^{-14}$) ratios for the present-day solar photosphere^{23,24}. For a cross-section ratio

$\sigma(^6\text{Li})/\sigma(^{10}\text{Be})$ lying between 30 and 90 (refs 23–26), the corresponding spallation solar (hereafter referred to as spall-S) $^6\text{Li}/\text{H}$ would range between 3×10^{-13} and 9×10^{-13} , corresponding to a spall-S ($^6\text{Li} + ^7\text{Li}$)/H ratio in the range $(9\text{--}27) \times 10^{-13}$. Scaling the present day solar Li/H at 1.45×10^{-11} , this spall-S Li would represent between 6% and 19% of the total solar Li. Mass balance calculations (Fig. 3) show that a $^7\text{Li}/^6\text{Li}$ ratio of 31 for the photosphere can be obtained by mixing 6–19% of spall-S Li with solar Li whose isotopic ratio should be ≥ 150 in the burning zone. This isotopic ratio of 150 for Li in the burning zone would be in agreement with any destruction ratio of ^6Li relative to ^7Li higher than 1.5 (Fig. 3).

Although Li production in the Sun by energetic flares ($> 10 \text{ MeV}$)

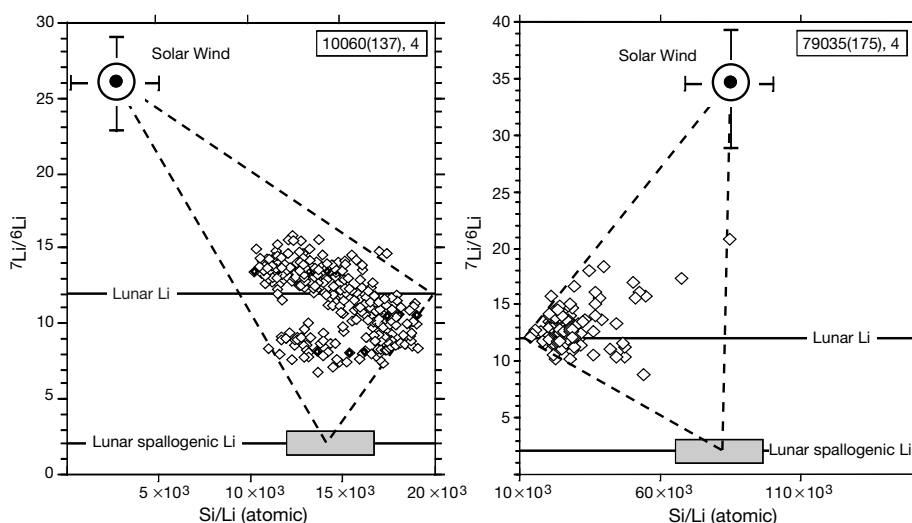


Figure 2 The $^7\text{Li}/^6\text{Li}$ isotopic ratio as a function of the atomic Si/Li ratio in lunar soils 10060 and 79035. In such a diagram, mixing processes produce linear relationships (the triangle defined by the dashed lines) between the lunar ($^7\text{Li}/^6\text{Li} = 12.14$), the spallogenic ($^7\text{Li}/^6\text{Li} = 2.0$; ref. 12) and the solar-wind (SW) components. The Si/Li ratio of the soil layers enriched in SW Li is highly variable, either higher or lower than the lunar Si/Li ratio.

Note that despite a lack of SW Li at the beginning of the profile (see profile in Fig. 1), sample 10060 shows clear mixing relationships between the three components. This might be explained by an experimental sampling effect due to the presence in the sputtered volume of a larger number of grains with highly variable solar-wind concentrations²⁷.

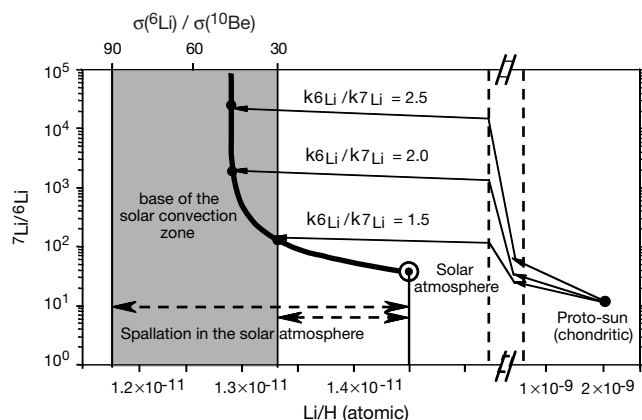


Figure 3 Evolution in the Sun of the ${}^7\text{Li}/{}^6\text{Li}$ ratio versus the Li/H ratio. All the calculations are scaled to $\text{Li}/\text{H} = 1.45 \times 10^{-11}$ and ${}^7\text{Li}/{}^6\text{Li} = 31$ for the solar atmosphere. The solid mixing curve shows the compositions at the base of the convection zone compatible with (1) the observed composition of the solar atmosphere and (2) ${}^7\text{Li}/{}^6\text{Li} = 2$ for Li produced by spallation. The shaded area indicates the Li/H ratio at the base of the solar convection zone compatible with the measured excess of ${}^{10}\text{Be}$ in lunar soils²³ (taking spallation cross-section ratios $30 < [\sigma({}^6\text{Li})/\sigma({}^{10}\text{Be})] < 90$). In order to satisfy the two constraints derived from the ${}^7\text{Li}/{}^6\text{Li}$ and ${}^{10}\text{Be}/\text{H}$ ratios of the solar atmosphere, the deep convection zone should have ${}^7\text{Li}/{}^6\text{Li} \geq 150$. Assuming a first-order rate for Li destruction, the ${}^7\text{Li}/{}^6\text{Li}$ evolution from the chondritic to the convection-zone value is also shown; this indicates that any destruction-rate ratio $k_{6\text{Li}}/k_{7\text{Li}} \geq 1.5$ is compatible with the ${}^7\text{Li}/{}^6\text{Li}$ ratio (see text).

seems the most likely interpretation of our results, it is important to note that the excess ${}^7\text{Li}$ detected in lunar soils is a solar-wind component ($< 10\text{ keV}$) because it is implanted at depths $< 30\text{ nm}$ (Fig. 1). This implies that the production of the spallogenic components by the flares must be decoupled from their escape to space. The minimum time required to reach the observed $[{}^6\text{Li}/\text{H}]_{\text{spall-S}}$ ratio can be calculated, assuming for simplicity that all ${}^6\text{Li}$ is produced by spallation ($[{}^6\text{Li}/\text{H}]_{\text{spall-S}} = 4.5 \times 10^{-13} = 1.45 \times 10^{-11}/(1 + 31)$) and that no Li destruction occurs at depth in the Sun. We calculate a value of 245 years from the ratio $[{}^6\text{Li}/\text{H}]_{\text{spall-S}}/\Phi_{6\text{Li}}$. The production rate ($\Phi_{6\text{Li}}$) of ${}^6\text{Li}$ is $\Phi_{\text{H}}\sigma({}^6\text{Li})[{}^{16}\text{O} + {}^{12}\text{C}]$, with Φ_{H} the solar proton flux at high energy at the surface of the Sun ($3.2 \times 10^6\text{ cm}^{-2}\text{ s}$; ref. 25), $\sigma({}^6\text{Li})$ the spallation cross-section for ${}^6\text{Li}$ on ${}^{16}\text{O}$ and ${}^{12}\text{C}$ (15 mbarn; ref. 25) and $[{}^{16}\text{O} + {}^{12}\text{C}]$ the relative solar abundances of carbon and oxygen (${}^{16}\text{O}/\text{H}$ and ${}^{12}\text{C}/\text{H} = 8.5 \times 10^{-4}$ and 3.63×10^{-4} , respectively²). According to this calculation, products of spallation reactions accumulate in the Sun's atmosphere for hundreds of years at least. □

Received 4 May; accepted 28 September 1999.

- Reeves, H. in *Origin and Evolution of the Elements* (eds Prantzos, N., Vangioni-Flam, E. & Cassé, M.) 168–197 (Cambridge Univ. Press, 1993).
- Anders, E. & Grevesse, N. Abundances of the elements: meteoritic and Solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Herbig, G. H. Lithium abundances in F5–G8 dwarfs. *Astrophys. J.* **141**, 588–609 (1965).
- Vauclair, S., Vauclair, G., Schatzman, E. & Michaud, G. Hydrodynamical instabilities in the envelopes of main-sequence stars: constraints implied by the lithium, beryllium and boron observations. *Astrophys. J.* **223**, 567–582 (1978).
- Brun, A. S. & Turck-Chieze, S. in *LiBeB, Cosmic Rays, and Related X- and Gamma Rays* (eds Ramaty, R., Vangioni-Flam, E., Cassé, M. & Olive, K.) 64–72 (ASP Conf. Ser., Vol. 171, Astronomical Society of the Pacific, San Francisco, 1999).
- Delbourgo-Salvador, P., Malinie, G. & Audouze, J. in *Isotopic Ratios in the Solar System* 267–272 (Toulouse Cepadues-Editions, 1984).
- Chaussidon, M. & Robert, F. ${}^7\text{Li}/{}^6\text{Li}$ and ${}^{10}\text{B}/{}^{11}\text{B}$ variations in chondrules from the Semarkona unequilibrated chondrite. *Earth Planet. Sci. Lett.* **164**, 577–589 (1998).
- Borg, J., Dran, J. C., Durrieu, L., Jouret, C. & Maudret, M. High voltage electron microscope studies of fossil nuclear particle tracks in extraterrestrial matter. *Earth Planet. Sci. Lett.* **8**, 379–386 (1970).
- Geiss, J. Solar wind composition and implications about the history of the Solar system. *Proc. 13th Int. Cosmic Ray Conf.* Vol. 5, 3375–3398 (Univ. of Denver, 1973).
- Becker, R. H. & Clayton, R. N. Nitrogen abundances and isotopic compositions in lunar samples. *Proc. Lunar Planet. Sci. Conf.* **6**, 2131–2149 (1975).
- Deliyannis, C. P., Demarque, P. & Kawaler, S. D. Lithium in halo stars from standard evolution. *Astrophys. J. Suppl.* **73**, 21–65 (1990).

- Eugster, O. & Bernas, R. Li, B, Mg and Ti isotopic abundances and search for trapped Solar wind Li in Apollo 11 and 12 material. *Proc. Lunar Planet. Sci. Conf.* **2**, 1461–1469 (1971).
- Ritzenhoff, S., Schröter, E. H. & Schmidt, W. The lithium abundance in sunspots. *Astron. Astrophys.* **328**, 695–701 (1997).
- Bibring, J. P. et al. Ultrathin amorphous coatings on lunar dust grains. *Science* **175**, 753–755 (1972).
- Keller, L. P. & McKay, D. S. The nature and origin of rims on lunar soil grains. *Geochim. Cosmochim. Acta* **61**, 2331–2341 (1997).
- Borg, J. et al. A Monte Carlo model for the exposure history of lunar dust grains in the ancient Solar wind. *Earth Planet. Sci. Lett.* **29**, 161–174 (1976).
- Reedy, R. O. in *The Ancient Sun* (eds Pepin, R. O., Eddy, J. A. & Merrill, R. B.) 365–386 (GCA Suppl. Vol. 13, Lunar and Planetary Institute, Houston, Texas) (Pergamon, 1979).
- Christensen-Dalsgaard, J., Gough, D. O. & Thompson, M. J. The depth of the Solar convection zone. *Astrophys. J.* **378**, 413–437 (1991).
- Audouze, J., Boulade, O., Malinie, G. & Poilane, Y. Galactic evolution of the lithium isotopes. *Astron. Astrophys.* **127**, 164–168 (1983).
- Balachandran, S. C. & Bell, R. A. Shallow mixing in the solar photosphere inferred from revised beryllium abundances. *Nature* **392**, 791–793 (1998).
- Chaussidon, M. & Robert, F. Nucleosynthesis of ${}^{11}\text{B}$ -rich boron in the pre-solar cloud recorded in meteoritic chondrules. *Nature* **374**, 337–339 (1995).
- Smith, V. V., Lambert, D. L. & Nissen, P. E. The ${}^7\text{Li}/{}^6\text{Li}$ ratio in the metal-poor halo dwarfs HD 19445 and HD 84937. *Astrophys. J.* **408**, 262–276 (1993).
- Nishiizumi, K., Caffee, M. W. & Arnold, J. R. ${}^{10}\text{Be}$ from the active Sun. *Proc. Lunar Planet. Sci. Conf.* **28**, 1027–1028 (1997).
- Jull, A. J. T., Lal, D. & Donahue, D. J. Evidence for a non-cosmogenic implanted ${}^{14}\text{C}$ component in lunar samples. *Earth Planet. Sci. Lett.* **136**, 693–702 (1995).
- Sisterson, J. M., Kim, K., Caffee, M. W. & Reedy, R. C. ${}^{10}\text{Be}$ and ${}^{26}\text{Al}$ production in lunar rock 68815: revised production rates using new cross section measurements. *Proc. Lunar Planet. Sci. Conf.* **28**, 1327–1328 (1997).
- Reeves, H. On the origin of the light elements. *Annu. Rev. Astron. Astrophys.* **12**, 437–467 (1974).
- Hashizume, K., Marty, B. & Wieler, R. Single grain analyses of the nitrogen isotopic composition in the lunar regolith: in search of the Solar wind component. *Proc. Lunar Planet. Sci. Conf.* **30**, 1567–1568 (1999).

Acknowledgements

We thank J. Audouze, S. Vauclair and R. Reedy for discussions, and G. Huss for contributions to this Letter. This work was supported by INSU-CNRS, CNES and MNHN.

Correspondence and requests for materials should be addressed to M.C.
(e-mail: chocho@crpg.cnrs-nancy.fr).

Carbon nanotube intramolecular junctions

Zhen Yao*, Henk W. Ch. Postma*, Leon Balents† & Cees Dekker*

* Department of Applied Sciences and DIMES, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

† Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

The ultimate device miniaturization would be to use individual molecules as functional devices. Single-wall carbon nanotubes (SWNTs) are promising candidates for achieving this: depending on their diameter and chirality, they are either one-dimensional metals or semiconductors^{1,2}. Single-electron transistors employing metallic nanotubes^{3,4} and field-effect transistors employing semiconducting nanotubes⁵ have been demonstrated. Intramolecular devices have also been proposed which should display a range of other device functions^{6–11}. For example, by introducing a pentagon and a heptagon into the hexagonal carbon lattice, two tube segments with different atomic and electronic structures can be seamlessly fused together to create intramolecular metal-metal, metal-semiconductor, or semiconductor-semiconductor junctions. Here we report electrical transport measurements on SWNTs with intramolecular junctions. We find that a metal-semiconductor junction behaves like a rectifying diode with nonlinear transport characteristics that are strongly asymmetric with respect to bias polarity. In the case of a metal-metal junction, the conductance appears to be strongly suppressed and it displays a power-law dependence on temperatures and

applied voltage, consistent with tunnelling between the ends of two Luttinger liquids. Our results emphasize the need to consider screening and electron interactions when designing and modelling molecular devices. Realization of carbon-based molecular electronics will require future efforts in the controlled production of these intramolecular nanotube junctions.

Figure 1a and b display examples of atomic force microscope images of nanotube junction devices. Each nanotube consists of two straight segments, both with measured heights of ~ 1 nm, connected by a sharp kink of about 40° . In principle, such kinks can originate from two different mechanisms: a pentagon–heptagon (5–7) topological defect pair as illustrated in Fig. 1c, or a local mechanical deformation in a uniform nanotube. Both experiments and simulations^{12,13} have shown that a nanotube will elastically deform under a small bending stress, and buckle if the local curvature exceeds a critical value. In this case, a height increase at the buckling point is expected. From our microscope images, however, the kinks appear to be very abrupt, and there seems to be no height increase at the kinks. This suggests that the kinks correspond to the 5–7 defects. Insertion of a pentagon (or heptagon) into the hexagonal carbon network creates a caplike (or saddlelike) curvature. A 5–7 pair will in general lead to a kink (see Fig. 1c). In order to generate a kink of a large angle, the pentagon and heptagon must be placed on the opposite sides of the kink⁷. Further structural evidence for the occurrence of 5–7 pairs comes from a double-kink sample shown in the Supplementary Information.

The nanotube shown in Fig. 1a is lying on three electrodes. The upper straight segment has a two-probe linear resistance of $110\text{ k}\Omega$ at room temperature (inset Fig. 2a) with no gate-voltage dependence, indicating that it is metallic⁵. In striking contrast, the current–voltage (I – V) characteristic across the kink (shown in Fig. 2a) is highly nonlinear and asymmetric, resembling that of a

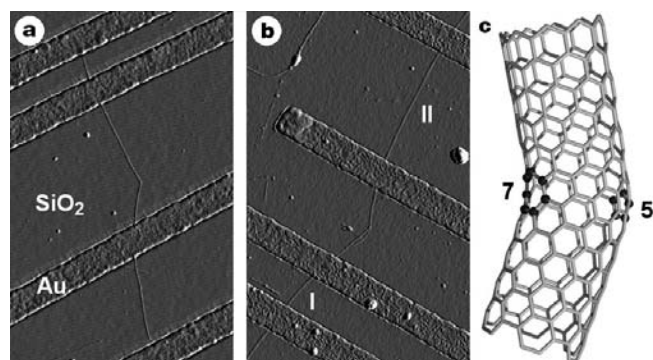


Figure 1 Tapping-mode atomic force microscope amplitude images of examples of nanotube junction devices. **a**, **b**, Nanotubes that contain a single kink of 36° and 41° respectively. We have detected in total four single-kink and one double-kink (see Supplementary Information) samples out of the ~ 500 devices we have inspected. These kinks can be associated with pentagon and heptagon defects which join two nanotube pieces with different diameters and chiralities. **c**, Illustration of the carbon-bond network of a kink junction constructed between an 'armchair' tube and a 'zigzag' tube, where 5 denotes a pentagon, 7 denotes a heptagon, and the atoms in the pentagon and heptagon are highlighted by dark balls. The nanotube in **a** is on top of three electrodes. Device **b**, with four leads connected, represents the ideal geometry, where both straight segments, I and II, and the kink can be characterized separately. The 250-nm wide, 20-nm-thick titanium–gold electrodes are embedded in SiO_2 with a height difference of less than 1 nm, which minimises the deformation of the nanotubes. This is achieved by standard electron-beam lithography, evaporation and lift-off, combined with an additional step of reactive-ion etching. A highly-doped silicon substrate below SiO_2 is used as a gate to vary the carrier density of the nanotubes. Nanotubes are deposited on top of the electrodes by spin coating of a drop of SWNT suspension in dichloroethane. The two-probe resistance of an individual straight metallic SWNT is typically $\leq 100\text{ k}\Omega$ at room temperature, more than an order of magnitude less than in earlier experiments^{3,5} (Z.Y. *et al.*, unpublished results). The low tube–electrode contact resistance proves essential in this work.

rectifying diode. The current rises sharply above a threshold voltage for a positive bias applied to the upper electrode. There is a small increase in current for large reverse bias. At around zero bias, the junction impedance is immeasurably high. Rectifying behaviour as a possible signature for metal–semiconductor (M–S) heterojunctions was previously observed in a film of entangled SWNTs (ref. 14). However, the junctions were not explicitly observed, and contact with metal electrodes and intertube tunnelling, which can also give rise to the observed behaviour, cannot be excluded. The I – V characteristics presented here are clearly associated with the kink defect. We have also measured I – V characteristics for different gate voltages V_g (Fig. 2b). The I – V curve seems to shift along the voltage axis as a function of V_g (inset Fig. 2b). Upon application of a negative V_g , the I – V curve becomes increasingly asymmetric. The curve at $V_g = -4\text{ V}$ shows almost perfect rectifying behaviour up to the largest bias voltage (-4 V) that we have applied. The strong gate modulation demonstrates unambiguously that the lower nanotube segment is semiconducting and that the kink is an M–S heterojunction.

For an isolated nanotube containing an M–S junction, the Fermi energy of the metallic segment is aligned with the middle of the energy gap of the semiconducting segment. The alignment is modified when the sample is connected to metal electrodes.

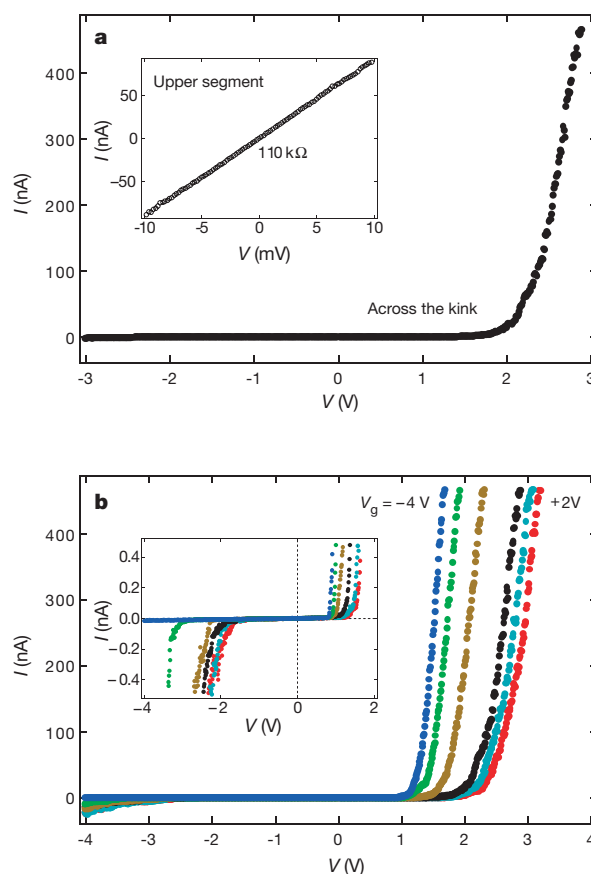


Figure 2 Current–voltage characteristics across the metal–semiconductor junction of Fig. 1a, showing rectifying behaviour. The data are taken at 100 K. The results at room temperature are similar, but the data are noisier. Inset to **a**, the I – V curve for the upper straight segment measured at room temperature. The low resistance value and the absence of a gate effect indicate that this segment is metallic. In **a**, the gate is grounded. In **b**, the gate voltages from right to left are 2 V, 1 V, 0 V, -1 V , -2 V and -4 V respectively. Inset to **b**, expanded view of the small-current region which shows more clearly the onset of conduction for both bias polarities. The junction resistance around zero bias is $>250\text{ G}\Omega$. The strong gate modulation demonstrates convincingly that the kink is a metal–semiconductor heterojunction.

Owing to the higher work function for the metal, both pieces will be hole-doped by the electrodes^{1,5}. The difference in electronic structures and screening properties of metallic and semiconducting nanotubes will give rise to different band-bending profiles in the tube segments away from the electrodes and subsequently a Schottky-type barrier at the M–S interface, which may explain the rectifying behaviour across the junction. Further modelling is required to understand the details of the I – V characteristics.

Each of the two straight segments of the sample in Fig. 1b lies on two electrodes, which permits their separate characterization and also two- and four-probe measurements across the kink. Figure 3 shows the temperature (T) dependence of the two-probe conductance (G) measured between different electrodes at zero bias. At room temperature, the resistances of the two straight segments are 56 and 101 k Ω respectively, with no gate-voltage dependence, demonstrating that both are metallic. The resistance across the junction is 608 k Ω , which is much higher. As the temperature decreases, all the conductances decrease monotonically. The conductance across the junction is much more temperature-dependent than that of the two straight segments, and decreases over one order of magnitude as the temperature decreases from 300 K to 50 K. Four-probe and two-probe measurements across the kink give essentially the same conductance value, indicating that the observed temperature dependence is completely dominated by the junction itself.

Although a defect is always expected to degrade the conductance, simple tight-binding calculations, which neglect electron–electron interactions, of the conductance across a metal–metal (M–M) kink junction cannot produce the large suppression seen in our experiments¹⁵. In Fig. 3 the data are plotted on a double-logarithmic scale. It appears that all the conductances can be fitted with a power-law function of T , $G \propto T^\alpha$, which is particularly convincing for the conductance across the junction because of its strong temperature dependence. Power-law behaviour of G versus T was reported recently for metallic ropes of nanotubes¹⁶, and was interpreted as a signature for electron–electron correlations^{17,18}. It has been known for decades that electron interactions are of great importance in one-dimensional transport¹⁹. Electrons form a correlated ground state called Luttinger liquid (LL), in which the tunnelling density of

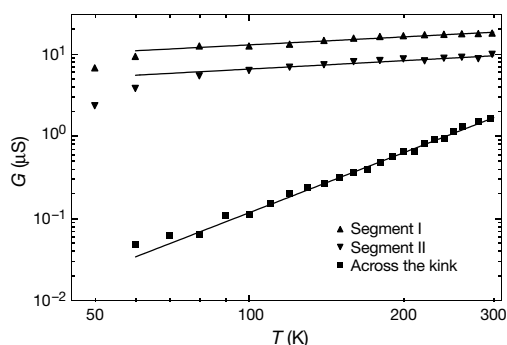


Figure 3 Linear-response two-probe conductances G of segments I and II and across the metal–metal junction of Fig. 1b, plotted against temperature T on a double-logarithmic scale. The data are fitted (solid lines) by the power law, $G(T) \propto T^\alpha$, which is associated with the suppression of tunnelling density of states in a Luttinger liquid. The exponents α for the two straight segments are 0.34 and 0.35 respectively. The fit is particularly convincing for the data across the kink. An exponent of 2.2 is obtained, which is consistent with end-to-end tunnelling between two Luttinger liquids. A thermally activated form for transport over a tunnel junction of barrier height U , $G \propto \exp(-U/k_B T)$, does not fit well. For the two straight segments, the overall behaviour is similar to that reported for ropes of nanotubes¹⁶, and is typical for our samples of individual SWNTs with similar conductance values (H. P. *et al.*, unpublished results). The low-temperature deviation from the power law is due to the Coulomb-blockade effect, which sets in when $k_B T$ becomes comparable to the charging energy, $E_c = e^2/2C$, needed to put an extra electron onto the nanotube, where C is the total capacitance of the tube.

states is suppressed as a power-law function of energy, $\rho(E) \propto E^\alpha$. Tunnelling into the end of an LL is more strongly suppressed than into the bulk, that is, the exponent for end tunnelling α_{end} is larger than the exponent for bulk tunnelling α_{bulk} . For an LL connected to three-dimensional reservoirs by tunnel barriers, the tunnelling conductance in the linear-response regime should vary as $G(T) \propto T^\alpha$ (for $eV \ll k_B T$, where e is the electron charge and k_B is the Boltzmann constant) and the differential conductance dI/dV at large bias ($eV \gg k_B T$) should vary as $dI/dV \propto V^\alpha$. An interesting question arises when a tunnel junction is placed between two LLs (ref. 19). To a first-order approximation, the tunnelling conductance

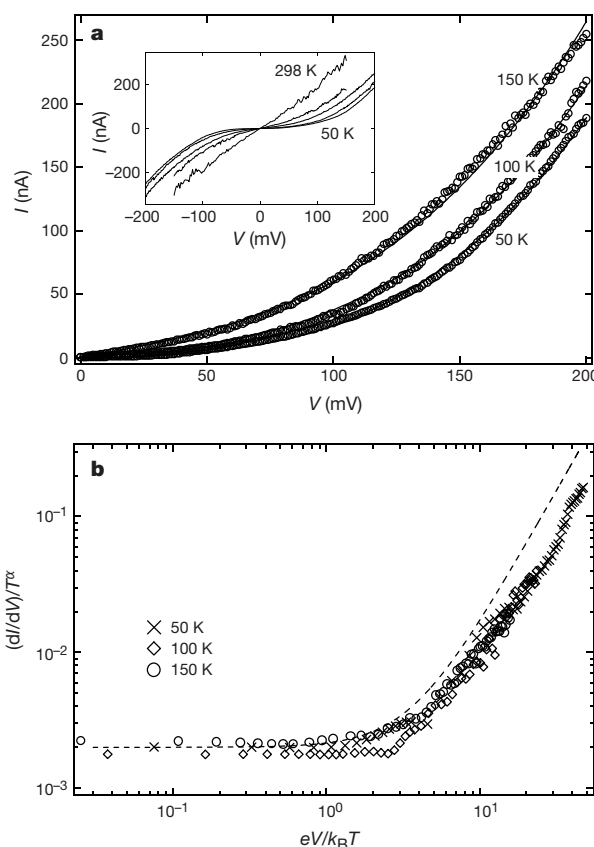


Figure 4 Large-bias transport characteristics measured across the metal–metal junction of Fig. 1b. **a**, Nonlinear I – V characteristics at different temperatures showing consistency with the Luttinger model. The data are fitted (solid lines) by a phenomenological functional form $I = C_1 T^\alpha V(1 + C_2(eV/k_B T)^\alpha)$, where C_1 and C_2 are constants. An exponent of $\alpha = 2.1$ is obtained which is similar to the value from G versus T measurement across the kink. The inset shows I – V curves at temperatures of 298 K, 200 K, 150 K, 100 K and 50 K. **b**, Scaled differential conductance $(dI/dV)/T^\alpha$ across the junction plotted against $eV/k_B T$, where $\alpha = 2.2$ is obtained from the power-law fit of G versus T across the kink. The data taken at various temperatures appear to collapse onto one curve. dI/dV is obtained by numerically differentiating the I – V curves shown in **a**. The dashed line represents the theoretical scaling function expected for the end-to-end tunnelling between two Luttinger liquids. At large bias, the measured conductance is lower than the theoretical prediction. Several factors may be responsible for the deviation. We have assumed that the voltage applied between the two contacts drops entirely across the junction, which is reasonable at small bias. As bias voltage is increased, however, more and more voltage drop will occur at the contacts because tunnelling conductance across the contacts (bulk tunnelling) increases less strongly as a function of voltage than that across the junction (end-to-end tunnelling). This effectively leads to reduced conductance across the junction compared to the case where one assumes a complete voltage drop across the junction. Additionally, the tunnelling amplitude across the junction may be energy-dependent at large bias. If the junction is viewed as a nanometer-sized capacitor, charging it to a few hundred millivolts may lead to a strong electrostatic force which could possibly deform the junction and reduce the transmission.

across the junction is proportional to the product of the end-tunnelling density of states on both sides and therefore still varies as a power law of energy, but with an exponent twice as large as the end tunnelling, that is, $\alpha_{\text{end-end}} = 2\alpha_{\text{end}}$.

Single-wall carbon nanotubes are truly one-dimensional conductors and are thus expected to behave as LLs. The strength of electron interactions is described by the so-called Luttinger parameter g . For non-interacting electrons $g = 1$, whereas for repulsive Coulomb interactions $g < 1$. The exponents α for tunnelling into the bulk and end of a nanotube LL are related to g as $\alpha_{\text{bulk}} = (g^{-1} + g - 2)/8$ and $\alpha_{\text{end}} = (g^{-1} - 1)/4$ respectively¹⁸. In our experiments, bulk tunnelling is expected for the measurements of the straight segments. Fitting the high-temperature data for the two straight segments yields similar exponents $\alpha = 0.34$ and 0.35 , which gives a value of $g \approx 0.22$ using the above expression for bulk tunnelling. This g value agrees with theoretical estimates^{17,18} and with that reported for ropes¹⁶ and clearly indicates the strong electron-electron interactions in SWNTs. Having established that both straight segments are LLs, we expect that the conduction process between the middle two contacts takes place via end-to-end tunnelling between the two LLs as well as tunnelling in and out of the two LLs through the contacts. However, the end-to-end tunnelling dominates the energy dependence of the process. Substituting the above g value into $\alpha_{\text{end-end}} = (g^{-1} - 1)/2$, an exponent of 1.8 is expected for the tunnelling conductance across the junction. Fitting the experimental data with a power law yields $\alpha = 2.2$, which is indeed close to this value. This analysis is not based on any presumed g value (which is sample-dependent). The self-consistency of bulk and end-to-end tunnelling thus provides strong evidence for the Luttinger model.

We have also measured large-bias I - V characteristics across the M-M junction at different temperatures, which provide an independent verification of the LL theory. As is evident from Fig. 4a, the I - V curves are nonlinear at all temperatures. The current is expected to increase as $V^{\alpha+1}$ for $eV \gg k_B T$. We consider the differential conductance which, assuming a constant tunnelling amplitude across the junction, is predicted to satisfy a scaling function:

$$\frac{dI}{dV} \propto T^\alpha \sinh\left(\frac{eV}{2k_B T}\right) \left| \Gamma\left(1 + \frac{\alpha}{2} + i \frac{eV}{2\pi k_B T}\right) \right|^2 \times \left\{ \frac{1}{2} \coth\left(\frac{eV}{2k_B T}\right) - \frac{1}{\pi} \text{Im} \left[\Psi\left(1 + \frac{\alpha}{2} + i \frac{eV}{2\pi k_B T}\right) \right] \right\}$$

Here α is the end-to-end tunnelling exponent and Γ and Ψ are the gamma and digamma functions respectively. For $\alpha = 2$, a good approximation (exact when $\alpha = 2$) of the above expression is given by $dI/dV \propto T^\alpha (1 + 3(eV/2\pi k_B T)^\alpha)$, which clearly shows the expected power-law behaviour at large bias. The expression suggests that if we scale dI/dV by T^α and V by T , then curves obtained at different temperatures should all collapse onto one universal curve. This is shown in Fig. 4b for our experimental data. The scaled conductance is constant for small V but crosses over to a power law for $eV/k_B T > 3$. Its behaviour is qualitatively similar to the theoretical scaling function (dashed line) in Fig. 4b. The deviation at large bias may be explained by a reduced voltage drop and/or a reduced tunnelling amplitude across the kink (see Fig. 4 legend). Overall, however, end-to-end tunnelling between two LLs seems to capture the main physics of the observed data. □

Received 24 June; accepted 29 September 1999.

- Wildöer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1998).
- Odum, T. W., Huang, J., Kim, P. & Lieber, C. M. Atomic structure and electronic properties of single-walled carbon nanotubes. *Nature* **391**, 62–64 (1998).
- Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).

- Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
- Tans, S. J., Verschueren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
- Chico, L., Crespi, V. H., Benedict, L. X., Louie, S. G. & Cohen, M. L. Pure carbon nanoscale devices: Nanotube heterojunctions. *Phys. Rev. Lett.* **76**, 971–974 (1996).
- Lambin, Ph., Fonseca, A., Vigneron, J. P., Nagy, J. B. & Lucas, A. A. Structural and electronic properties of bent carbon nanotubes. *Chem. Phys. Lett.* **245**, 85–89 (1995).
- Saito, R., Dresselhaus, G. & Dresselhaus, M. S. Tunnelling conductance of connected carbon nanotubes. *Phys. Rev. B* **53**, 2044–2050 (1996).
- Charlier, J.-C., Ebbesen, T. W. & Lambin, Ph. Structural and electronic properties of pentagon-heptagon pair defects in carbon nanotubes. *Phys. Rev. B* **53**, 11108–11113 (1996).
- Menon, M. & Srivastava, D. Carbon nanotube “T junctions”: Nanoscale metal-semiconductor-metal contact devices. *Phys. Rev. Lett.* **79**, 4453–4456 (1997).
- Chico, L., López Sancho, M. P. & Muñoz, M. C. Carbon-nanotube-based quantum dot. *Phys. Rev. Lett.* **81**, 1278–1281 (1998).
- Iijima, S., Brabec, C. J., Maiti, A. & Bernholc, J. Structural flexibility of carbon nanotubes. *J. Chem. Phys.* **104**, 2089–2092 (1996).
- Yakobson, B. I., Brabec, C. J. & Bernholc, J. Nanomechanics of carbon tubes: Instabilities beyond the linear response. *Phys. Rev. Lett.* **76**, 2511–2514 (1996).
- Collins, P. G., Zettl, A., Bando, H., Thess, A. & Smalley, R. E. Nanotube nanodevice. *Science* **278**, 100–103 (1997).
- Chico, L., Benedict, L. X., Louie, S. G. & Cohen, M. L. Quantum conductance of carbon nanotubes with defects. *Phys. Rev. B* **54**, 2600–2606 (1996).
- Bockrath, M. *et al.* Luttinger-liquid behaviour in carbon nanotubes. *Nature* **397**, 598–601 (1999).
- Egger, R. & Gogolin, A. Effective low-energy theory for correlated carbon nanotubes. *Phys. Rev. Lett.* **79**, 5082–5085 (1997).
- Kane, C., Balents, L. & Fisher, M. P. A. Coulomb interactions and mesoscopic effects in carbon nanotubes. *Phys. Rev. Lett.* **79**, 5086–5089 (1997).
- Fisher, M. P. A. & Glazman, L. I. in *Mesoscopic Electron Transport* (eds Kouwenhoven, L. P., Sohn, L. L. & Schön, G.) 331–373 (Kluwer Academic, Boston, 1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. E. Smalley and co-workers for providing the indispensable single-wall carbon nanotube materials, M. P. Anantram, S. J. Tans and A. A. Odintsov for helpful discussions, V. Meunier for the atomic coordinates used in Fig. 1c, and M. de Jonge and A. van den Ende for experimental assistance. The work was supported by the Dutch Foundation for Fundamental Research on Matter (FOM).

Correspondence and requests for materials should be addressed to C.D. (e-mail: dekker@qt.tn.tudelft.nl).

Design and synthesis of an exceptionally stable and highly porous metal-organic framework

Hailian Li*, Mohamed Eddaoudi†, M. O'Keeffe* & O. M. Yaghi†

Materials Design and Discovery Group, * Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA

† Department of Chemistry, University of Michigan, 930 North University, Ann Arbor, Michigan 48109-1055, USA

Open metal-organic frameworks are widely regarded as promising materials for applications^{1–15} in catalysis, separation, gas storage and molecular recognition. Compared to conventionally used microporous inorganic materials such as zeolites, these organic structures have the potential for more flexible rational design, through control of the architecture and functionalization of the pores. So far, the inability of these open frameworks to support permanent porosity and to avoid collapsing in the absence of guest molecules, such as solvents, has hindered further progress in the field^{14,15}. Here we report the synthesis of a metal-organic framework which remains crystalline, as evidenced by X-ray single-crystal analyses, and stable when fully desolvated and when heated up to 300 °C. This synthesis is achieved by borrowing ideas from metal carboxylate cluster chemistry, where an organic

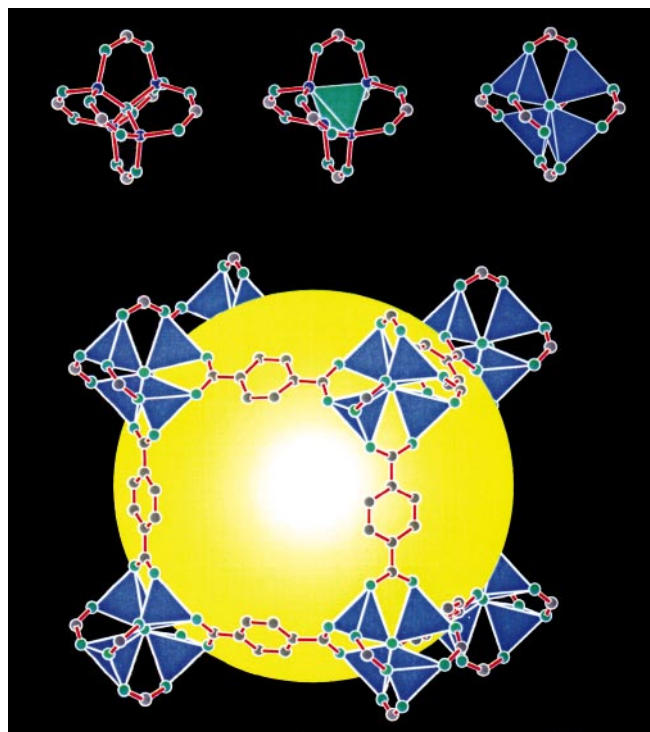


Figure 1 Construction of the MOF-5 framework. Top, the $\text{Zn}_4(\text{O})_{12}\text{C}_6$ cluster. Left, as a ball and stick model (Zn, blue; O, green; C, grey). Middle, the same with the $\text{Zn}_4(\text{O})$ tetrahedron indicated in green. Right, the same but now with the ZnO_4 tetrahedra

indicated in blue. Bottom, one of the cavities in the $\text{Zn}_4(\text{O})(\text{BDC})_3$, MOF-5, framework. Eight clusters (only seven visible) constitute a unit cell and enclose a large cavity, indicated by a yellow sphere of diameter 18.5 Å in contact with 72 C atoms (grey).

dicarboxylate linker is used in a reaction that gives supertetrahedron clusters when capped with monocarboxylates. The rigid and divergent character of the added linker allows the articulation of the clusters into a three-dimensional framework resulting in a structure with higher apparent surface area and pore volume than most porous crystalline zeolites. This simple and potentially universal design strategy is currently being pursued in the synthesis of new phases and composites, and for gas-storage applications.

The tetranuclear supertetrahedral cluster motif shown in Fig. 1a is adopted by a number of metal carboxylates (acetate, benzoate and pivalate), where combination of Zn^{2+} and the appropriate carboxylic acid yields the oxide centred cluster as a distinct and well-defined unit¹⁶. Working towards an extended network based on this cluster, we viewed its Zn–O–C motif as a secondary building unit capable of assembly under similar reaction conditions but in the presence of a dicarboxylate instead of a monocarboxylate. Our earlier work¹⁸ involving the copolymerization of 1,4-benzenedicarboxylate (BDC) with metal ions pointed to its rigid, planar, and divergent attributes, which, when coupled with the rigidity of this secondary building unit, would be ideally suited in forming the target extended three-dimensional framework shown in Fig. 1b.

We found that diffusion of triethylamine into a solution of zinc (II) nitrate and H_2BDC in N,N' -dimethylformamide (DMF)/chlorobenzene resulted in the deprotonation of H_2BDC and its reaction with Zn^{2+} ions. A small amount of hydrogen peroxide was added to the reaction mixture in order to facilitate the formation of O^{2-} expected at the centre of the secondary building unit. Elemental microanalysis of the resulting colourless cubic crystals suggested that their formula was $\text{Zn}_4\text{O}(\text{BDC})_3 \cdot (\text{DMF})_8(\text{C}_6\text{H}_5\text{Cl})$; here we call this material MOF-5, where MOF indicates metal organic framework. The presence of DMF and chlorobenzene guests was confirmed by solid-state ^{13}C NMR, which showed sharp characteristic peaks at the chemical shifts expected for the carbon atom resonances of these guests. Ion mass spectrometry, performed on the liberated guest vapours as the material was heated from 22 to 350 °C, showed their respective parent-ion masses.

An X-ray single-crystal diffraction study on the as-synthesized materials reveals the expected topology (Fig. 2). The structure may be derived from a simple cubic six-connected net in two stages: first, the nodes (vertices) of the net are replaced by clusters of secondary building units; second, the links (edges) of the net are replaced by finite rods ('struts') of BDC molecules. The core of the cluster

Table 1 Gas and liquid vapour sorption in desolvated MOF-5

Sorbate	<i>T</i> (°C)	Amount sorbed (mg g ⁻¹)	Sorbate molecules per unit cell	Free volume (cm ³ g ⁻¹)	Free volume (cm ³ cm ⁻³)
Ar	-194	1,492	230	1.03	0.61
N ₂	-194	831	183	1.04	0.61
CH ₂ Cl ₂	22	1,211	88	0.93	0.55
CHCl ₃	22	1,367	71	0.94	0.55
C ₆ H ₆	22	802	63	0.94	0.55
CCl ₄	22	1,472	59	0.94	0.56
C ₆ H ₁₂	22	703	51	0.92	0.54

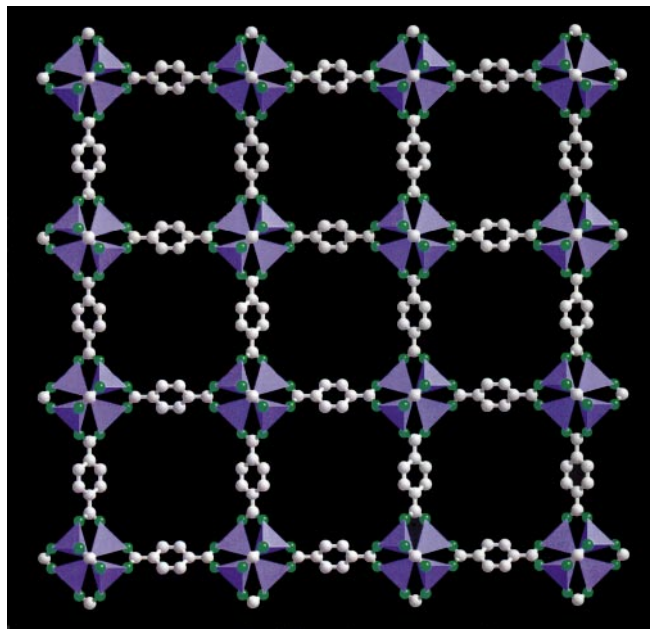


Figure 2 Representation of a {100} layer of the MOF-5 framework shown along the a -axis (C, grey; O, green). The Zn_4O tetrahedra are indicated in purple. Properties of MOF-5 are as follows. Single crystals of as-synthesized $\text{Zn}_4\text{O}(\text{BDC})_3 \cdot (\text{DMF})_8(\text{C}_6\text{H}_5\text{Cl})$, MOF-5, are at 213 ± 2 K cubic, space group $Fm-3m$ with $a = 25.6690(3)$ Å, $V = 16,913.2(3)$ Å³, and $Z = 8$. Analysis for MOF-5. (1) Elemental microanalysis. Calculated (%): C, 44.21; H, 5.02; N, 7.64; Zn, 17.82%. Found (%): C, 43.25; H, 5.29; N, 7.56; Zn, 17.04%. (2) Solid-state ^{13}C NMR. First, CP MAS. Guests $(\text{CH}_3)_2\text{NC}(\text{O})\text{H}$: $(\text{CH}_3)_2$ -, δ (p.p.m.) 37.81 and 32.71; $-\text{C}(\text{O})-$, δ 164.93. $\text{C}_6\text{H}_5\text{Cl}$: δ 132.04, 130.41, 128.76. Framework, $-\text{COO}$, δ 175.93; $-\text{C}(\text{COO})$, δ 139.11; $o\text{-C}_6\text{H}_4$, δ 130.41. Second, static. Guests $(\text{CH}_3)_2\text{NC}(\text{O})\text{H}$: $(\text{CH}_3)_2$ -, δ 37.63 and 32.59; $-\text{C}(\text{O})-$, δ 164.67. $\text{C}_6\text{H}_5\text{Cl}$: δ 130.22 (broad). Framework, no peaks observed. (3) Mass spectrometry, 22–350 °C. $(\text{CH}_3)_2\text{NC}(\text{O})\text{H}$: M^+ , 73 a.m.u. ($\text{C}_3\text{H}_7\text{NO}^+$).

consists of a single O atom bonded to four Zn atoms, forming a regular Zn_4O tetrahedron. Each edge of each Zn tetrahedron is then capped by a $-\text{CO}_2$ group to form a $\text{Zn}_4(\text{O})(\text{CO}_2)_6$ cluster. We note that a related zinc phosphate structure exists¹⁷, in which the C atoms are replaced by the P atoms of PO_4 tetrahedra, which serves to link the clusters together. The important step taken in the present work is the replacement of the $\text{O}_2\text{--P--O}_2$ phosphate links by $\text{O}_2\text{--C--}$

$\text{C}_6\text{H}_5\text{Cl}$: M^+ , 77 (C_6H_5^+); M^+ , 112 a.m.u. ($\text{C}_6\text{H}_5^{35}\text{Cl}^+$) and 114 a.m.u. ($\text{C}_6\text{H}_5^{37}\text{Cl}^+$) in 3:1 intensity ratio. Properties of single crystals of fully desolvated MOF-5, $\text{Zn}_4\text{O}(\text{BDC})_3$, are at 169 ± 2 K cubic, space group $Fm-3m$ with $a = 25.8849(3)$ Å, $V = 17,343.6(8)$ Å³, and $Z = 8$ (for empirical formula, $\text{C}_{24}\text{H}_{12}\text{O}_{13}\text{Zn}_4$), R , $R_w = 0.023$, 0.026. Analysis for $\text{Zn}_4\text{O}(\text{BDC})_3$. Elemental microanalysis. Calculated (%): C, 37.45; H, 1.57; N, 0.00%. Found (%): C, 36.94; H, 1.66; N, 0.26%. Solid-state ^{13}C NMR (CP MAS and static) showed no resonances due to any guests. Mass spectrometry (22–350 °C) showed no peaks, indicative of absence of any guests in the pores. Thermogravimetric analysis showed no weight loss up to 410 °C. Single crystals of $\text{Zn}_4\text{O}(\text{BDC})_3$, fully desolvated and heated at 300 °C, are at 149 ± 2 K cubic, space group $Fm-3m$ with $a = 25.8496(3)$ Å.

$\text{C}_6\text{H}_4\text{--C--O}_2$ (BDC) ‘struts’, to give an extended network having a three-dimensional intersecting channel system with 12.94-Å spacing between the centres of adjacent clusters.

The framework atoms in MOF-5 take up only a small fraction of the available space in the crystal. If overlapping spheres with van der Waals radii (1.5, 1.2, 1.7, 1.5 Å for Zn, H, C and O, respectively) are placed at the atomic positions, the space not so occupied is 80% of

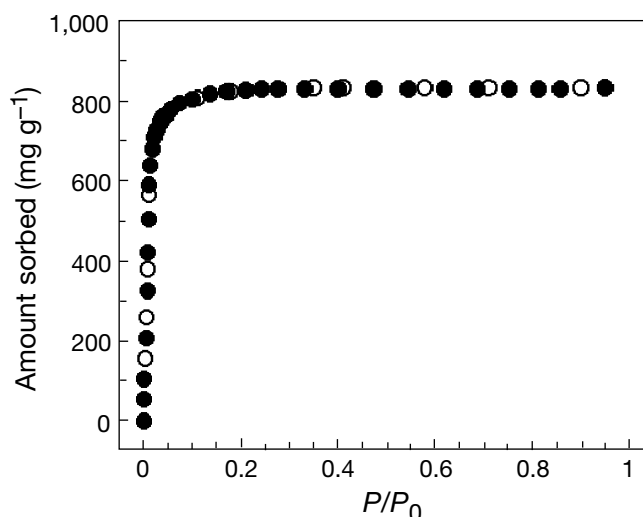


Figure 3 Nitrogen gas sorption isotherm at 78 K for MOF-5 (filled circles, sorption; open circles desorption). P/P_0 is the ratio of gas pressure (P) to saturation pressure (P_0), with $P_0 = 746$ torr.

the crystal volume. The non-framework space is divided into two types of cavity lined with C and H atoms. The centre of the smaller of these has 24 H atoms at 7.10 Å and 24 C atoms at 7.97 Å; the centre of the larger (Fig. 1) has 72 C atoms at 9.26 Å and 48 H atoms at 9.47 Å. The aperture joining the two cavities has 8 H at 5.10 Å and 8 C at 5.70 Å from its centre. Allowing for the van der Waals radii, spheres of material with diameter 15.1 Å and 11.0 Å could fit in the large and small cavities, respectively, and the aperture would admit the passage of a sphere of diameter 8.0 Å. Assuming the same molar volume as in the respective liquids, we find that in the original material the guest molecules— $8 \times (8 \text{ DMF} + 1 \text{ C}_6\text{H}_5\text{Cl})$ —have a volume of $9,512 \text{ Å}^3$, which is 55% of the unit cell volume. Indeed, we show below that 55–61% of the space is accessible to guest species.

Due to their high mobility, the guests can be fully exchanged with chloroform. More significant is that all the chloroform guests can be evacuated from the pores within 3 h at room temperature and 5×10^{-5} torr without loss of framework periodicity. In fact, the desolvated single crystals were checked by elemental microanalysis, ^{13}C NMR, mass spectrometry, and thermal gravimetry to confirm the absence of any guests, while optical microscopy was used to show that they maintained their transparency, morphology and crystallinity. This has allowed us to perform another X-ray single-crystal diffraction study on the fully desolvated form of this framework—a study that revealed mono-crystallinity with cell parameters and atomic positions coincident with those of the as-synthesized material. In fact, the highest peaks and lowest valleys in the ΔF map are 0.25 and -0.17 electrons per Å^3 scattered randomly both near the framework and the void space. These contrast with the relatively large values (1.56 and -0.45 electrons per Å^3) found in the as-synthesized crystals.

Further evidence for the stability of the framework was obtained by heating the fully desolvated crystals in air at 300°C for 24 h, which had no effect on either their morphology or crystallinity as evidence by another X-ray single-crystal diffraction study. Here again, the cell parameters obtained were unaltered relative to those found for the unheated desolvated crystals, illustrating the rigidity and stability of the framework in the absence of guest molecules. These results are interesting as the density (0.59 g cm^{-3}) of the desolvated crystals is among the lowest recorded for any crystalline material.

To evaluate the pore volume and the apparent surface area of this framework, the gas and vapour sorption isotherms for the desolvated sample were measured using an electromicrogravimetric balance (Cahn 1000) set-up and an already published procedure¹⁸. The sorption of nitrogen revealed a reversible type I isotherm (Fig. 3), with concordant results obtained with large crystals (0.2 mm) and microcrystals (30 μm). The same sorption behaviour was observed for argon and organic vapours such as CH_2Cl_2 , CHCl_3 , CCl_4 , C_6H_6 and C_6H_{12} . Similar to those of most zeolites, the isotherms are reversible and show no hysteresis upon desorption of gases from the pores. Using the Dubinin–Radushkevich equation, pore volumes of $0.61\text{--}0.54 \text{ cm}^3 \text{ cm}^{-3}$ were calculated and found to be nearly identical for all adsorbates shown (Table 1), which indicates the presence of uniform pores. Assuming a monolayer coverage of N_2 (which may not be strictly correct with such large cavities) the apparent Langmuir surface area was estimated at $2,900 \text{ m}^2 \text{ g}^{-1}$. Zeolites, which generally have higher molar masses than $\text{Zn}_4\text{O}(\text{BDC})_3$, have pore volumes that range from $0.18 \text{ cm}^3 \text{ cm}^{-3}$ for analcime to $0.47 \text{ cm}^3 \text{ cm}^{-3}$ for zeolite A¹⁹. □

Received 20 May; accepted 17 September 1999.

1. Kinoshita, Y., Matsubara, I., Higuchi, T. & Saito, Y. The crystal structure of bis(adiponitrile) copper(I) nitrate. *Bull. Chem. Soc. Jpn* **32**, 1221–1226 (1959).
2. Kitagawa, S., Munakata, M. & Tanimura, T. Tetranuclear copper(I)-based infinite one-dimensional chain complex. *Chem. Lett.* 623–626 (1991).
3. Abrahams, B. F., Hoskins, B. F. & Robson, R. Infinite polymeric frameworks consisting of three dimensionally linked rod-like segments. *J. Am. Chem. Soc.* **113**, 3606–3607 (1991).

4. Zaworotko, M. J. Cooperative bonding affords a wholesome story. *Nature* **386**, 220–221 (1997).
5. Fagan, P. J. & Ward, M. D. Building molecular crystals. *Sci. Am.* **267**, 48–54 (1992).
6. Carlucci, L., Ciani, G., Proserpio, D. M. & Sironi, A. Interpenetrating diamondoid frameworks of silver(I) cations linked by N,N'-bidentate molecular rods. *J. Chem. Soc. Chem. Commun.* 2755–2756 (1994).
7. Mallouk, T. E. Crowns get organized. *Nature* **387**, 350–351 (1997).
8. Gardner, G. B., Venkataraman, D., Moore, J. S. & Lee, S. Spontaneous assembly of a hinged coordination network. *Nature* **374**, 792–795 (1995).
9. Yaghi, O. M., Li, G. & Li, H. Selective binding and removal of guests in a microporous metal-organic framework. *Nature* **378**, 703–706 (1995).
10. Lu, J. *et al.* Coordination polymers of $\text{Co}(\text{NCS})_2$ with pyrazine and 4,4'-bipyridine: syntheses and structures. *Inorg. Chem.* **36**, 923–928 (1997).
11. Vaid, T. P., Lobkovsky, E. B. & Wolczanski, P. T. Covalent 3- and 2-dimensional titanium-quinone networks. *J. Am. Chem. Soc.* **119**, 8742–8743 (1997).
12. Liu, F. Q. & Tilley, T. D. A coordination network based on d^0 transition-metal centers: synthesis and structure of the [2,4]-connected layered compound $[(\text{TiCl}_4)_2\text{Si}(\text{C}_6\text{H}_4\text{CN}-p)_4] \cdot 1.5\text{C}_2\text{H}_8$. *J. Chem. Soc. Chem. Commun.* 103–104 (1998).
13. MacGillivray, L. R., Groeneman, R. H. & Atwood, J. L. Design and self-assembly of cavity-containing rectangular grids. *J. Am. Chem. Soc.* **120**, 2676–2677 (1998).
14. Yaghi, O. M., Li, H., Davis, C., Richardson, D. & Groy, T. L. Synthetic strategies, structure patterns, and emerging properties in the chemistry of modular porous solids. *Acc. Chem. Res.* **31**, 575–585 (1998).
15. Kepert, C. J. & Rosseinsky, M. J. Zeolite-like crystal structure of an empty microporous framework. *J. Chem. Soc. Chem. Commun.* 375–376 (1999).
16. Clegg, W. *et al.* Crystal structures of three basic zinc carboxylates together with infrared and FAB mass spectrometry studies in solution. *Inorg. Chim. Acta* **186**, 51–60 (1991).
17. Harrison, W. T. A. *et al.* Synthesis and characterization of a new family of thermally stable open framework zincophosphate/arsenate phases: $\text{M}_2\text{Zn}_4\text{O}(\text{XO}_4)_3 \cdot n\text{H}_2\text{O}$ ($\text{M} = \text{Na}, \text{K}, \text{Rb}, \text{Li}, \dots$; $\text{X} = \text{P}, \text{As}$; $n = 3\text{--}5$). Crystal structures of $\text{Rb}_2\text{Zn}_4\text{O}(\text{PO}_4)_3 \cdot 3.5\text{H}_2\text{O}$, $\text{K}_2\text{Zn}_4\text{O}(\text{AsO}_4)_3 \cdot 4\text{H}_2\text{O}$, and $\text{Na}_2\text{Zn}_4\text{O}(\text{PO}_4)_3 \cdot 6\text{H}_2\text{O}$. *Chem. Mater.* **8**, 691–700 (1996).
18. Li, H., Eddaoudi, M., Groy, T. L. & Yaghi, O. M. Establishing microporosity in open metal-organic frameworks: gas sorption isotherms for $\text{Zn}(\text{BDC})$ ($\text{BDC} = 1,4\text{-benzenedicarboxylate}$). *J. Am. Chem. Soc.* **120**, 8571–8672 (1998).
19. Breck, D. W. *Zeolite Molecular Sieves* (Wiley & Sons, New York, 1974).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank F. Hollander and R. Staples for X-ray structure determinations. This work was supported by the National Science Foundation (M.O.K. and O.M.Y.), the Department of Energy (O.M.Y.) and Nalco Chemical Company.

Correspondence and requests for materials should be addressed to O.M.Y. at the University of Michigan (e-mail: oyaghi@umich.edu).

Abrupt termination of the 1997–98 El Niño in response to a Madden–Julian oscillation

Yukari N. Takayabu*, Toshio Iguchi†, Misako Kachi‡, Akira Shibata‡ & Hiroshi Kanzawa*

* National Institute for Environmental Studies, Tsukuba, Ibaraki, 305-0053, Japan

† Communications Research Laboratory, Tokyo, 184-8795, Japan

‡ Earth Observation Research Center/NASDA, Tokyo, 106-0032, Japan

The role of the Madden–Julian oscillation—a global atmospheric wave in the tropics that is associated with convective activity and propagates eastwards with a period of about 30–60 days (refs 1, 2)—in triggering El Niño events has been discussed before^{3–8}. But its possible connection with a termination of El Niño has yet to be investigated, despite the difficulty in explaining the timing of El Niño terminations by the basic wind-induced oceanic-wave processes^{9,10}. For the extreme 1997–98 event, the mechanism of both onset and termination have been investigated³, but the reason for the abruptness of the termination has yet to be resolved. Here we present global data of precipitation, sea surface temperatures and wind speeds that show a precipitation system associated with an exceptionally strong Madden–Julian oscillation.

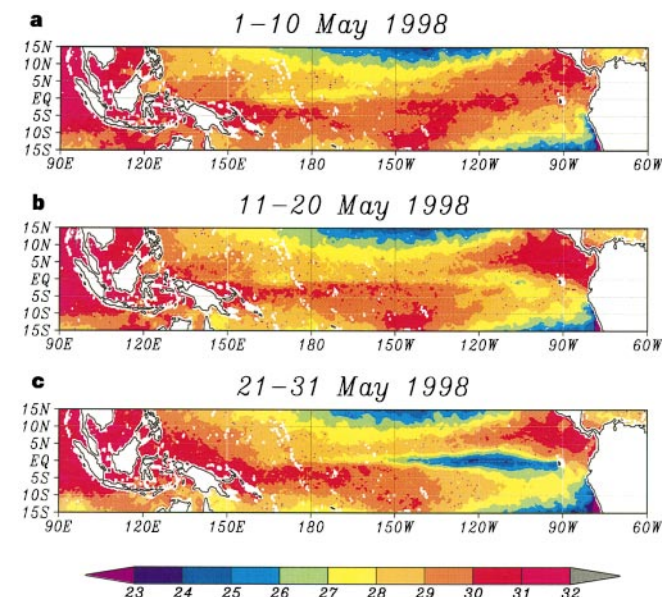


Figure 1 Time averaged sea surface temperature (SST) for 1–10, 11–20 and 21–31 May 1998. SST was retrieved from the TRMM Microwave Imager (TMI) data with the algorithm developed by Shibata¹². Units on the colour scale are °C.

tion travelling around the Equator in May 1998. The propagation of this atmospheric system was associated with an abrupt intensification of the easterly trade winds over the eastern equatorial Pacific Ocean. Combined with the already shallow equatorial thermocline in the central and eastern Pacific Ocean at that time³, these strong winds provided the triggering mechanism for the observed accelerated ending of the 1997–98 El Niño event.

Global precipitation data over the ocean were produced at $0.25^\circ \times 0.25^\circ$ longitude–latitude grids from Special Sensor Microwave/Imager (SSM/I) satellite data¹¹ by F. Wentz. The precipitation data obtained from the space-borne precipitation radar on the Tropical Rainfall Measuring Mission (TRMM) satellite were also utilized to interpolate the data over land. Daily averages of these precipitation data were constructed for the analysis. Daily sea surface temperature (SST) data were retrieved from the TRMM Microwave Imager data at $0.25^\circ \times 0.25^\circ$ longitude–latitude grids with an algorithm developed by Shibata *et al.*¹² Wind data were obtained from the global analysis data set produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). Twice-daily data at $2.5^\circ \times 2.5^\circ$ longitude–latitude grids were utilized.

The abrupt termination of the 1997–98 El Niño is shown in a sequence of time-averaged SST plots (Fig. 1). In the first ten-day period, the entire equatorial Pacific was still at the stage of El Niño and covered with the warm SST, except for a hint of the equatorial upwelling near the dateline which showed as a slightly colder SST patch. In the second ten days, the equatorial upwelling in the eastern Pacific started. In the last period, a robust ‘cold tongue’, or extended cold-water region, was established in the equatorial eastern Pacific.

Before these events, the thermocline in the central eastern Pacific Ocean had shallowed³, and thereby transferred the ocean into a state that allowed strong cooling of the SSTs by upwelling and mixing of subsurface water. This shallowing was first associated with equatorial waves emanating from the western boundary of the Pacific Ocean^{9,10}. In addition, it was enhanced by direct wind forcing by the southward shifts in equatorial westerly winds in the later stage of El Niño¹³.

A longitude–time section of the SST averaged for 3°N – 3°S is plotted in Fig. 2 for the period from 1 May to 1 June, with daily ECMWF surface wind vectors at the Equator. It is evident that a cold tongue of SST started to evolve from 110° – 130°W at around 13 May.

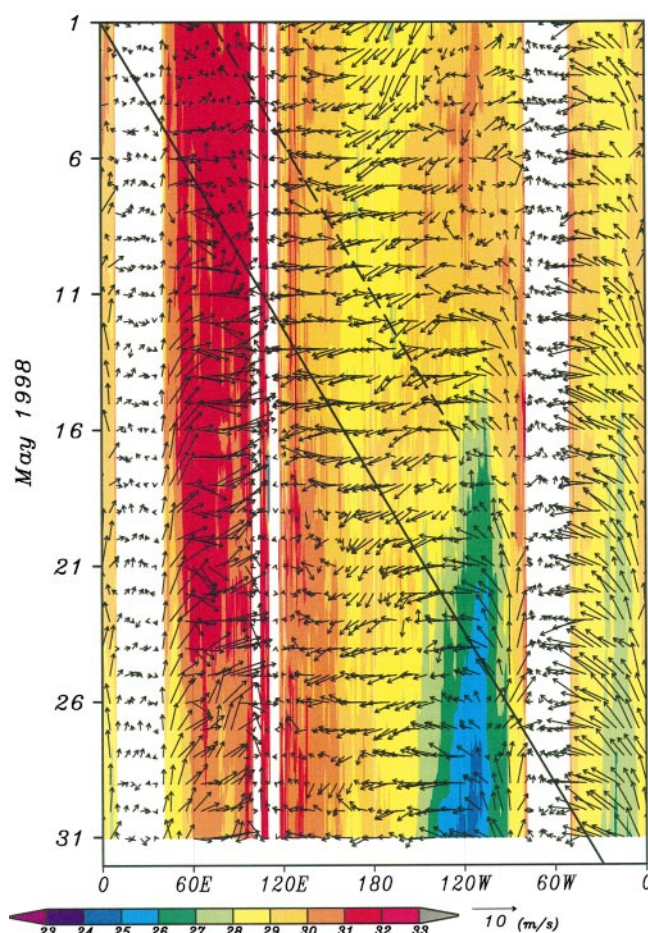


Figure 2 Time–longitude sections of SST and the surface winds. Colours indicate the SST derived from the TMI data and averaged from 3°N to 3°S . Units on the colour scale are °C. The vectors are the ECMWF surface winds on the Equator. The abscissa is the longitude starting from the Greenwich meridian, and the ordinate is for the period of 1 May–1 June 1998. The solid line corresponds to the propagation of the major rain system found in Fig. 3, and the dashed line shows that of the preceding rain system (see text for details).

An accompanying intensification of the easterly winds suggested by McPhaden³ is also shown by the ECMWF winds to the west of 90°W . Under the influence of the Earth’s rotation, the easterly winds on the Equator induce the equatorial oceanic upwelling that causes the SST cooling. Then, on 18–20 May, a sudden westward expansion of the cold tongue occurred, with a re-intensification of the easterlies to the west of 110°W . We note that these easterly-wind intensifications are part of eastward-propagating large-scale wind disturbances. We now consider what these propagating disturbances are.

The colours in Fig. 3 show a longitude–time section of SSM/I rain rate averaged for 10°N – 10°S . The rain data over land are interpolated with precipitation detected by the precipitation radar and depicted with grey shades. Eastward propagation of a global wavenumber-one precipitation system, which is identified as a Madden–Julian oscillation (MJO), is observed to complete the global equatorial circle in about 33 days, starting from near the Greenwich meridian on 1 May. The intensifications of the easterly winds are leading the propagation of the precipitation systems indicated in Fig. 3 by straight lines, especially in the region from 120°W to 100°W over the Pacific Ocean. The intensified trade winds accompanying the first weaker rain system, indicated with a dashed line, caused cold water to appear at the surface around 13 May. Then, the second intensification of the trade winds

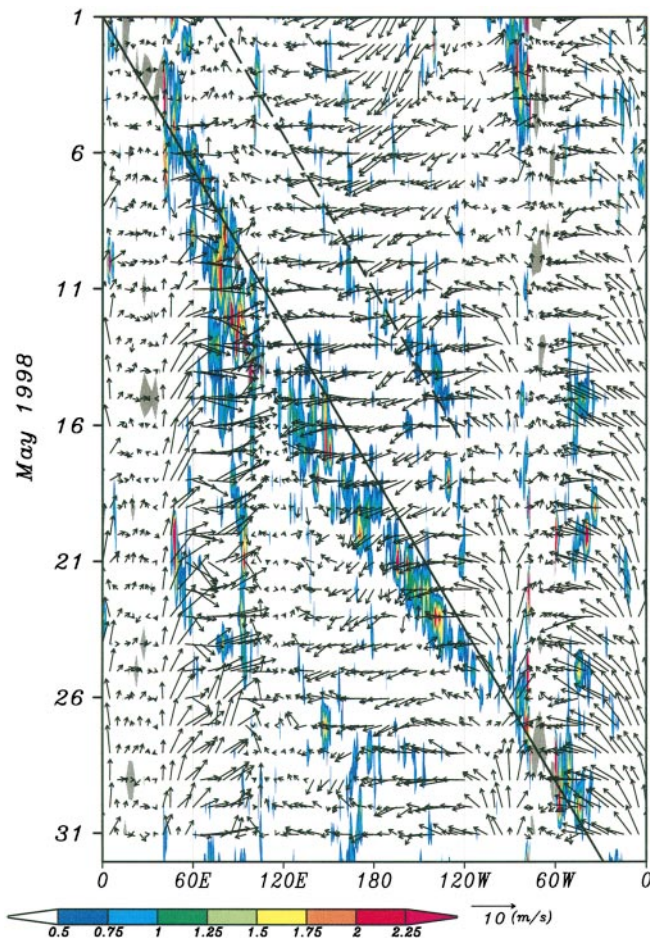


Figure 3 Time-longitude sections of the rain rate and the surface winds. Colours show the rain rate obtained from the SSM/I retrievals (see text), and the grey shades show that obtained from the TRMM precipitation radar with the PR2a25 algorithm (T.I., manuscript in preparation) for values $>0.5 \text{ mm h}^{-1}$, and averaged from 10° N to 10° S . Units on the colour scale are mm h^{-1} . The vectors are the ECMWF surface winds on the Equator.

accompanying the following major system, indicated with a solid line, caused the westward expansion of the cold tongue to start on 18–20 May.

A composite structure of this MJO is shown in Fig. 4. Base longitudes for each panel follow the propagation of the major precipitation system indicated by a solid line in Fig. 3. Figure 4a shows the composite SSM/I precipitation anomaly. A well organized precipitation system emerges near the composite centre. The primary rain area has a horseshoe shape straddling the Equator from about 5° N to 5° S , open to the east. Its longitudinal extent is about 60° . The preceding rain system that initiated the cold tongue is evident about 60° to the east of the composite centre.

The wind anomaly at the 200-hPa level (Fig. 4b) clearly exhibits the equatorial Rossby wave (an equatorial wave mode with the vortex structures off the Equator) response to the west of the precipitation centre and the Kelvin wave (an equatorial wave mode which has the wind components only parallel to the Equator) response to the east. This is the atmospheric response pattern to the heating imposed on the Equator, known as the ‘Matsuno–Gill pattern’^{14,15}. Here, the latent-heat release in the precipitation system works as the imposed forcing. In the lower troposphere at 700 hPa (Fig. 4c), on the other hand, the Rossby wave response diminishes and the Kelvin wave structure becomes dominant. This point will be discussed later.

The surface winds (Fig. 4d) show a conspicuous intensification of

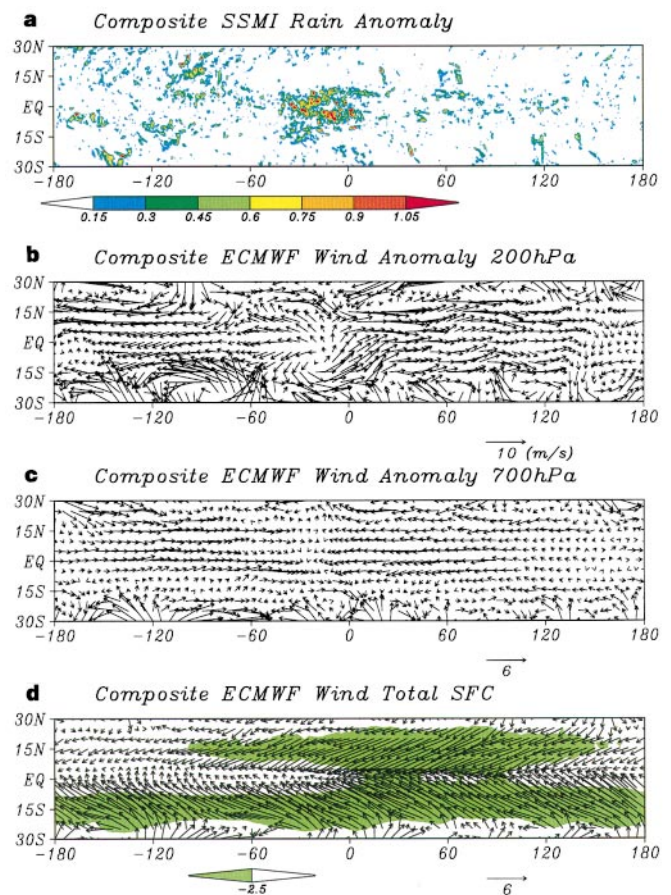


Figure 4 Composite SSM/I precipitation and ECMWF horizontal winds. Panel **a** is for precipitation, and composite winds are shown at 200 hPa (**b**), at 700 hPa (**c**) and at the surface (**d**). For **a–c**, anomalous fields from the zonal means are depicted. For the surface wind vectors, the total values are depicted, and the green colour shows where the easterly zonal wind speed exceeds 2.5 m s^{-1} . See text for details of the abscissa; the ordinate shows actual latitude.

the easterly components especially to the east of the precipitation centre. The wavenumber-one component is dominant here also. We can confirm the intensified easterly winds over the Equator, at around $+25^\circ$ and $+90^\circ$ relative longitudes, as indicated with green shading. Another significant feature at the surface is the strong wind convergence towards the Equator. This is the Ekman convergence associated with the easterly winds, under the effects of the surface friction and the rotation of the Earth. The convergence centres are along 5° N and 5° S , and have larger values ahead (to the east) of the precipitation system. As a result, belts rich in water vapour are produced ahead of the precipitation system, which helps to maintain its eastward propagation. Thus, an interaction between the precipitation processes and the dynamical waves maintained this MJO.

This MJO was most unusual, because its precipitation system travelled around the complete equatorial circle. Usually, precipitation systems associated with the MJOs are observed over the warm water pool region from the Indian Ocean to the central Pacific, but diminish over the eastern Pacific¹⁶. And only dry dynamical waves are considered to continue to propagate around the globe. We now consider why this MJO maintained its precipitation throughout its global trip.

Until early May, although the SST was warmer than 28° C over the entire equatorial Pacific Ocean, the cumulus convective activity was suppressed over the western Pacific Ocean. This was due to the

atmospheric subsidence associated with the Walker circulation, or the atmospheric large-scale zonal circulation. May 1998 corresponded to the transition period in terms of the Walker circulation: just after the large-scale subsidence over the western Pacific was relaxed, and just before the subsidence over the eastern Pacific was established. We consider that the relatively warm homogeneous SST distribution, without suppression of convection by large-scale subsidence, allowed the unusual global propagation of the MJO precipitation.

Another notable feature of this MJO was that it influenced the termination of the El Niño through the intensification of the easterly trade winds. A triggering effect of westerly wind bursts (associated with the MJO) on the initiation of El Niño events has been previously suggested^{3–8}. However, there have been no previous observations of the MJO that triggered the termination of El Niño events.

How an MJO influences the ocean depends on its dynamical structure. The MJO is usually considered to have the Rossby wave response of the 'Matsuno-Gill pattern' to the west of the convection centre. The intensification of the twin vortices from the Rossby wave response is favourable for the occurrence of westerly wind bursts. In our case, however, the lower-tropospheric disturbance lacked the Rossby wave response and had only the Kelvin wave response.

Wang and Xie¹⁷ studied theoretically the effects of mean wind shear on atmospheric equatorial waves. Their results indicated that the westerly wind shear (stronger westerly wind upward) traps the Rossby waves in the upper troposphere, while not much affecting the Kelvin waves. In our case, the mean zonal-wind difference between the 200-hPa and 850-hPa levels was 6.9 m s^{-1} , which is comparable to the 10 m s^{-1} that Wang and Xie assumed¹⁷.

Thus, the mean wind field not only allowed the global propagation of a convectively coupled MJO, but also affected its structure, suppressing the Rossby wave response in the lower troposphere. The resultant lower-tropospheric Kelvin wave disturbance superposed on the mean easterly winds in turn caused the intensification of the easterly trade winds, and brought about an unusual easterly-wind effect of the MJO on El Niño.

Slingo *et al.*¹⁸ suggested an important link between the MJO and the basic climate by comparing the performances of general circulation models. The May 1998 event re-emphasizes the significance of multi-scale interactions within the climate system, on which further studies are required. □

Received 4 June; accepted 20 September 1999.

- Madden, R. A. & Julian, P. R. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.* **28**, 702–708 (1971).
- Madden, R. A. & Julian, P. R. Description of global scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.* **29**, 1109–1120 (1972).
- McPhaden, M. J. Genesis and evolution of the 1997–98 El Niño. *Science* **283**, 950–954 (1999).
- Nitta, Ts. & Motoki, T. Abrupt enhancement of convective activity and low level westerly wind burst during the onset phase of 1986–87 El Niño. *J. Meteorol. Soc. Jpn* **65**, 497–506 (1987).
- Nitta, Ts. Development of a twin cyclone and westerly wind bursts during the initial phase of 1986–87 El Niño. *J. Meteorol. Soc. Jpn* **67**, 677–681 (1989).
- Nitta, Ts., Mizuno, T. & Takahashi, K. Multi-scale convective systems during the initial phase of 1986/87 El Niño. *J. Meteorol. Soc. Jpn* **70**, 317–335 (1992).
- Kindle, J. C. & Phoebus, P. A. The ocean response to operational westerly wind bursts. *J. Geophys. Res.* **100**, 4893–4920 (1995).
- Lau, K. M. & Chan, P. H. The 40–50 day oscillation and the El Niño/Southern Oscillation: A new perspective. *Bull. Am. Meteorol. Soc.* **67**, 533–534 (1986).
- Suarez, M. J. & Schopf, P. S. A delayed action oscillator for ENSO. *J. Atmos. Sci.* **45**, 3283–3287 (1988).
- Battisti, D. S. Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere-ocean model. *J. Atmos. Sci.* **45**, 2889–2919 (1988).
- Wentz, F. J. & Spencer, R. W. SSM/I rain retrievals within a unified all-weather ocean algorithm. *J. Atmos. Sci.* **55**, 1613–1627 (1998).
- Shibata, A., Imaoka, K., Kachi, M. & Murakami, H. SST observation by TRMM Microwave Imager aboard Tropical Rainfall Measuring Mission. *Umi no Kenkyu* **8**, 135–139 (1999). (In Japanese.)
- Harrison, D. E. & Vecchi, G. A. On the termination of El Niño. *Geophys. Res. Lett.* **26**, 1593–1596 (1999).
- Matsuno, T. Quasi-geostrophic motions in the equatorial area. *J. Meteorol. Soc. Jpn* **44**, 25–43 (1966).
- Gill, A. E. Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **106**, 447–462 (1980).
- Wang, B. & Rui, H. Synoptic climatology of transient tropical intraseasonal convection anomalies: 1975–1985. *Meteorol. Atmos. Phys.* **44**, 43–61 (1990).

- Wang, B. & Xie, X. Low-frequency equatorial waves in vertically sheared zonal flow. Part I: Stable waves. *J. Atmos. Sci.* **53**, 449–467 (1996).
- Slingo, J. M. *et al.* Intraseasonal oscillations in 15 atmospheric general circulation models: results from an AMIP diagnostic subproject. *Clim. Dyn.* **12**, 325–357 (1996).

Acknowledgements

We thank T. Nakazawa, Y. Suzuki, H. Fujinami, K. Imaoka and H. Murakami for help with processing the SSM/I and TRMM data.

Correspondence and requests for materials should be addressed to Y.N.T.

Combined dynamic and geochemical evidence for convergent melt flow beneath the East Pacific Rise

Marc Spiegelman* & Jennifer R. Reynolds*†

* Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA

Determining the flow of magma and solid mantle beneath mid-ocean ridges is crucial for understanding the dynamics of plate spreading and the formation of new oceanic crust. Theoretical models suggest a range of possible flow regimes—from passive, plate-driven flows^{1,2} to 'active', buoyantly driven solid convection^{3–7}—and have spurred an ambitious field programme to attempt to distinguish these flow fields using geophysical techniques⁸. Models that explore the geochemical consequences of melt transport⁹, however, suggest that these different flow fields can also have distinctive geochemical signatures. Here we compare model predictions to the chemistry of well located and closely sampled basalts from across the ridge-crest of the fast-spreading East Pacific Rise at 12° N (refs 10–12). These data show features that are not explained by traditional geochemical models of ocean-ridge magma generation, yet are consistent with the geochemical consequences of the new transport models that have passive mantle flow and convergent lateral melt migration. These results are also consistent with those of the seismic MELT experiment⁸, but add new information about the relative flow of melt and solid in the mantle which is probably unmeasurable by geophysical techniques.

The melt and solid flow fields for two end-member flow regimes proposed for mid-ocean ridges are shown in Fig. 1 (see ref. 9 for detailed discussion). In the regime shown in Fig. 1a, the solid mantle flow is driven by the spreading plates, producing a broad region of upwelling and melting (passive flow). The melt flow, however, is focused towards the ridge axis by pressure gradients induced by viscous stresses^{1,2}. In Fig. 1b, the solid mantle flow is driven both by the plates and by buoyancy forces induced by lateral variations in melt content (active flow). The buoyancy forces become significant at lower mantle viscosities, and drive small-scale mantle convection that promotes more rapid upwelling within a narrow region beneath the ridge axis^{3–7}. The lower viscous stresses, however, eliminate any significant convergent melt flow.

The outstanding question has been how to distinguish between these theoretical models using geological and/or geophysical observations. Both calculations in Fig. 1 form similar crustal thicknesses within a relatively narrow crustal accretion zone. Both calculations have the same maximum degree of melting and permeability functions, and produce mantle porosities that would probably be

† Present address: Monterey Bay Aquarium Research Institute, PO Box 628, 7700 Sandholt Rd, Moss Landing, California 95039, USA.

difficult to distinguish using physical techniques. Nevertheless, these flow calculations have distinct geochemical signatures which suggest that compositional variations in erupted magmas may be sensitive indicators of mantle dynamics⁹.

Figure 2 illustrates the geochemical signatures of these calculations for the simplest problem of trace elements with constant bulk partition coefficients and a uniform source composition. The flow models predict variations in both overall melt composition and in the spatial distribution of erupted compositions which can be significantly different from those calculated using models that neglect melt transport (see, for example, refs 13, 14). These variations are a direct consequence of the ability of melt and solid to travel, and mix, on different trajectories in two and three dimensions⁹. Convergent melt flow relative to solid flow (Fig. 1a) enriches the concentration of incompatible trace elements at the ridge axis by adding lower-degree melts from a wide region. Convergent solid flow within the partially molten region (Fig. 1b), however, produces axial melts that are extremely depleted in the most incompatible elements. These elements are removed by small amounts of melting and transport that occur off-axis before the solid undergoes most of its melting within the narrow upwelling zone. These two models also predict completely opposite signatures in the relative enrichment of incompatible trace elements in on-axis and off-axis melts (Fig. 2). The passive flow model produces magmas that are most enriched on-axis but become more depleted with distance off-axis within the crustal accretion zone (~5 km half-width in Fig. 1a). In contrast, the active flow model produces melts that become more enriched with distance off-axis, as do melting models that neglect transport.

The spatial trend of incompatible trace element enrichment or depletion is a robust consequence of the relative geometries of the melt and solid flow fields, and depends only on whether melt or solid flow is convergent. For this reason, we concentrate on com-

paring the composition of well located on- and off-axis lavas with similar sources. We also consider more quantitative details such as the amplitude of the signals and the width of the crustal accretion zone; however, these are more dependent on specific model parameters, and will be explored in depth elsewhere.

A detailed petrological and geological study around 12° N on the East Pacific Rise^{10–12} provides important data and interpretation that can be compared to predictions from the theoretical flow models. The study area (Fig. 3) covers 50 km of ridge axis, and extends 8 km west and 10 km east of the axis. Basaltic glasses from 143 rock core and dredge samples have been analysed by electron microprobe, and 52 samples (Fig. 3c) have also been analysed for major and trace elements by direct-current plasma spectrometry (DCP). Samples with full analyses form a 'training set' for classifying the remaining samples using a nearest-neighbour algorithm. Geological information is drawn from multi-beam bathymetry¹⁵, SeaMARC I side-scan sonar¹⁶ and camera tows. Further details are given elsewhere¹².

Within the "axial summit trough"¹⁷, one finds both normal mid-ocean-ridge basalts (N-MORBs) and small-volume eruptions of incompatible-trace-element enriched MORBs (T-MORBs) (Fig. 3), with remarkably few samples showing compositional evidence of simple mixing despite evidence for an axial magma chamber in this area¹⁸. These observations suggest that, at 12° N, individual melts

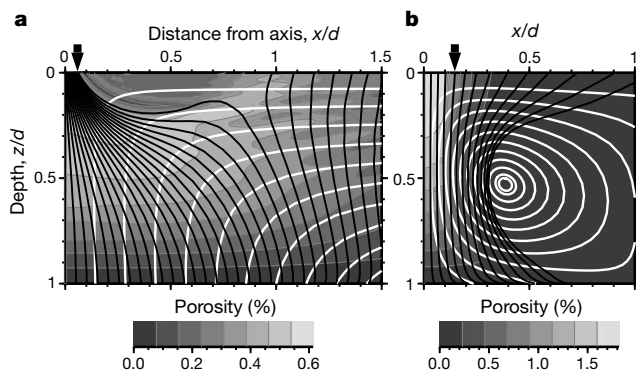


Figure 1 Two-dimensional calculations of melt and solid mantle flow beneath mid-ocean ridges for two end-member flow regimes. The diagrams are vertical cross-sections of the right half of the melting region. Horizontal (x) and vertical (z) distances are scaled to the depth where melting begins ($d \approx 80$ – 100 km). The black curves represent melt flow, while white curves represent solid mantle flow. The melting rate is proportional to the solid upwelling rate, and no melting occurs in downwelling regions or previously melted mantle. Both solutions have linear melting functions with maximum degree of melting $F_{\max} \approx 20\%$ and the same permeability function⁹. Arrows mark the width of the "crustal accretion region", defined as the zone over which ~85% of the melt produced within the box is added to the crust. If the crustal thickness produced over this region is h , then the depth scale is set by $d = h/0.075$ (that is, 6 km of crust implies $d = 80$ km). **a**, Passive, plate-driven flow. When solid mantle flow is dominated by plate motion, large pressure gradients due to shear in the viscous matrix can focus the flow of melt towards the spreading centre. **b**, 'Active' buoyancy driven flow. At lower mantle viscosities or slower plate velocities, variations in melt content drive more rapid upwelling beneath the spreading centre. The width of the upwelling and melting region narrows in compensation.

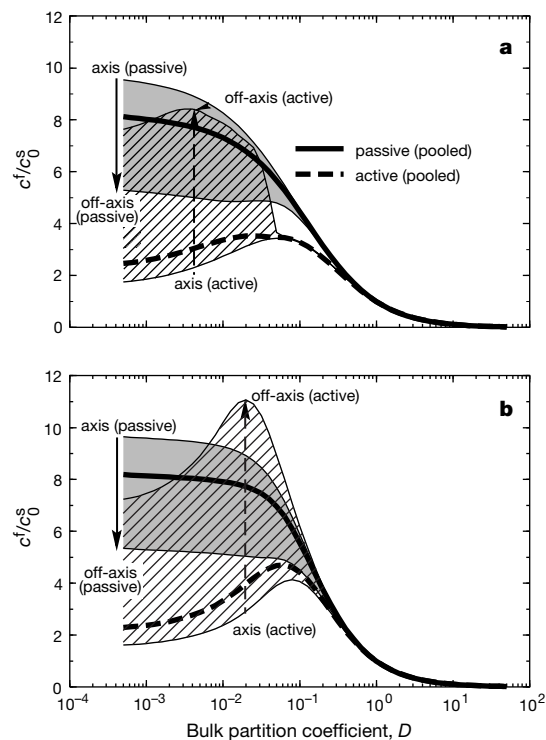


Figure 2 The geochemical signatures of the two flow models shown in Fig. 1. Results are shown for two end-member equilibration models of equilibrium transport (**a**) and disequilibrium transport (**b**), which are the open-system analogues of batch and fractional melting⁹. Each panel shows the range of melt compositions produced by the different models for trace elements with constant bulk partition coefficients. All melt concentrations (c') are normalized to the unmelted solid source (c_0). The shaded fields show results for the passive flow model, the hatched fields for the active model. The thin lines bounding these fields show the range of concentration patterns for melts 'erupted' between the axis ($x = 0$) and the off-axis edge of the accretion zone marked by arrows in Fig. 1 ($x = 0.06d$ for passive flow; $x = 0.14d$ for active flow). Heavy lines show the mean concentrations in the accretion zone, calculated as the sum of all the local melts weighted by their flux⁹. **a**, Equilibrium transport where melt and solid continuously re-equilibrate during transport. **b**, Disequilibrium transport which assumes fractional melting of the solid with mixing of incremental melts occurring along melt flow lines.

can preserve chemical differences through the crustal accretion process. In addition, Reynolds and Langmuir¹² have also identified a petrologically distinct lava type that only erupts off-axis (red symbols in Fig. 3a). In their original 1992 paper¹⁰, samples of this lava type were considered part of the temporal variation at the 12° N axis, and contributed scatter. When these samples are removed, the temporal variation in axial lava composition is even more systematic. The off-axis magma type is most readily identified in depleted compositions. Because of this, and to minimize the complicating effect of mantle source variations, we limit discussion here to N-MORB compositions and refer to the off-axis type as OA-N-MORB. The contrast between this lava type and the N-MORBs erupted on-axis form the basis for testing model predictions of melt and solid mantle flow.

Many samples of the OA-N-MORB type can clearly be identified with morphological features suggestive of off-axis eruption^{11,12} such as small seamounts (Fig. 3a), or flows and pillow mounds that cover abyssal hill faults (Fig. 3b). Observations during dives of the *Alvin* submersible at 9° 31' N and 9° 50' N have also documented small off-axis lava flows and volcanic constructions in that region^{17,19,20}. Of the 20 OA-N-MORB samples found at 12° N, 45% can be positively associated with features indicating off-axis eruption despite the incomplete side-scan coverage and limitations of mapping resolution. No sample of this composition type is found on the EPR axis here. A search of 123 earlier dredges along the EPR from 5° to 14° N identified seven OA-N-MORB samples off-axis elsewhere¹². No such samples were found along the axis, with a single exception from the intersection with the Clipperton transform.

For the samples with full major- and trace-element analyses, Fig. 4a compares the mean and standard deviation of the composition of N-MORB recovered within <1.5 km of the ridge axis, with

the mean and standard deviation of the off-axis-type lavas. The signature of low-pressure olivine–plagioclase–clinopyroxene fractionation has been removed by correcting the compositions to a constant 7.3% MgO, which is the mean MgO content of the full data set^{12,21}. Samples with MgO > 8.0% have been excluded because they may not lie on olivine–plagioclase–clinopyroxene fractionation trends. For clarity, all of the compositions are normalized by the mean of the axial N-MORB samples.

In terms of major elements, the OA-N-MORBs have signatures similar to the on-axis N-MORB, but tend to be less fractionated (higher MgO) on eruption, suggesting less cooling in the crust. The OA-N-MORBs are also characterized by unusually low abundances of trace elements that are incompatible during mantle melting (Ba to Sc) compared to N-MORB lavas from the ridge axis. This sense of off-axis depletion is the qualitative signature of models with convergent melt flow towards the ridge axis. This signal is clear in all incompatible trace elements with the possible exception of Sr which, while moderately incompatible during melting, behaves compatibly during low-pressure fractionation involving plagioclase. The variation in Sr content probably reflects more complex crustal fractionation processes than are accounted for by the corrections used here^{11,12}.

Although the differences in composition for the off-axis-type N-MORB are small (~20% less than the mean), they are also quantitatively consistent with the predictions of the models. Figures 4b and c compare the incompatible trace element data to the model calculations for passive flow, assuming a single bulk partition coefficient during melting and transport. These models are the same as those presented in ref. 9, and were calculated independently of the 12° N data. Nevertheless, the agreement between models and data is remarkably good, in both the overall shapes of the patterns

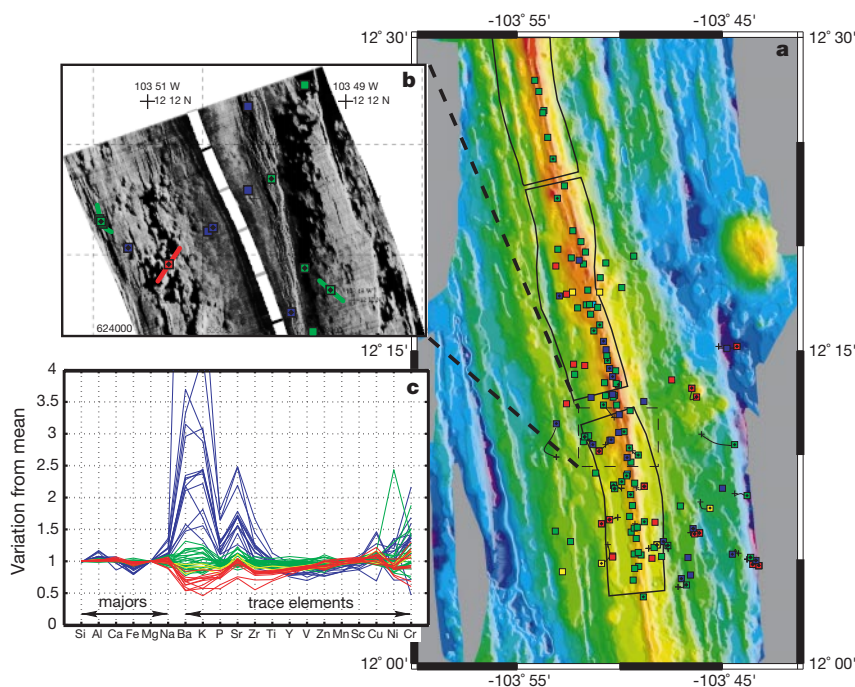


Figure 3 Map showing the location and composition of samples collected between 12° 00' and 12° 30' N on the East Pacific Rise^{10–12}. **a**, Bathymetric relief map of the region, illuminated from the east¹⁵. Sea-floor depths increase away from the ridge axis, from red (~2,600 m) to purple (~3,200 m). The thin black lines outline SeaMARC I sonar images¹⁶ that have been re-navigated. Dredges are shown by short heavy lines with squares on top. Rock core locations are shown by isolated squares. Samples with filled diamonds have full major and trace-element analyses (**c**) and are used to classify the remaining partial samples. Blue squares indicate incompatible trace-element enriched T-MORB; green squares indicate N-MORB; red squares indicate the off-axis-type N-MORB.

Yellow samples are intermediate between off-axis- and N-MORB. Details of sampling and analysis are given elsewhere^{10–12}. **b**, Close-up map showing an example of sample locations superimposed on Sea-MARC I imagery. The short red dredge (VE10) is an off-axis-type lava and was collected from hummocky extrusive flows that cover older sea-floor faults. Highly reflective regions are bright (white), and shadows are black. The axial summit trough is the fissured region in the centre of the eastern strip. **c**, Compositional variation of samples with full major- and trace-element analyses corrected to 7.3% MgO. The colour scheme is the same as the sample map. Compositions are all normalized to the mean composition of the N-MORB samples.

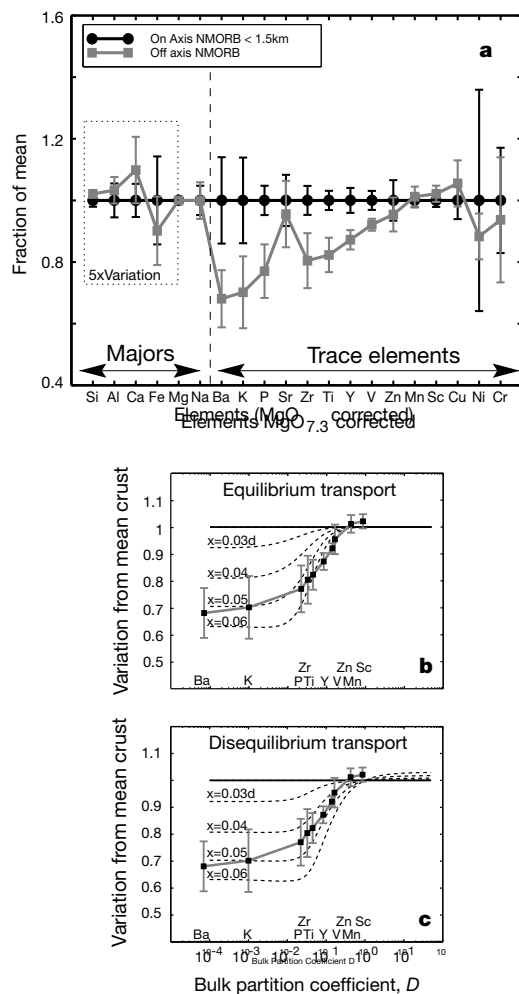


Figure 4 Comparison of on- and off-axis EPR magmas with model predictions. **a**, Major- and trace-element compositions of 12° N EPR basalt glasses with full analyses. Circles are the mean of 18 N-MORBs collected within <1.5 km of the axis, and the squares are the mean of 11 off-axis-type N-MORBs. The first six points are compositions of major-element oxides by microprobe, and the subsequent points are DCP analyses of trace elements shown roughly in order of compatibility. Variation from the mean in major elements for the first 5 elements is multiplied by a factor of 5. All values are fractionation-adjusted to 7.3% MgO and normalized to the mean of the on-axis N-MORB. Error bars show $\pm 1\sigma$ standard deviation which is a good proxy for the range of natural variation in each data group. **b**, Comparison of 12° N off-axis incompatible trace element patterns (squares) to model predictions for equilibrium melt transport in plate driven mantle flow with convergent melt migration (Fig. 1a). Compositions from the model are normalized to the model's mean "crustal" composition (thick lines in Fig. 2). Thin dashed lines indicate trace-element patterns calculated for various distances off-axis where the distance x is scaled to the depth of the melting region. The curve marked $x = 0.06d$ corresponds to the arrow in Fig. 1a, and would be 4.8 km for $d = 80$ km ($h = 6$ km). **c**, As **b** but for disequilibrium transport.

and the rough distances off-axis (of the order of several kilometres). Measured distances of samples, of course, should be used with caution in the absence of an independent estimate of the age of eruption. Analyses of U-series nuclides can provide information on eruption ages and crustal residence times up to ~ 300 kyr, and could provide an important independent test of these models in the future. Such measurements have already demonstrated clear off-axis eruptions at 9° 30' N^{17,19,22}.

The quantitative comparison of data and models is expected to change as more data and additional processes are included in the models, but the qualitative result is robust. The important point

is that the off-axis melts at 12° N appear more depleted in incompatible trace elements than the axial melts across a range of elements, consistent with convergent melt flow. Although there may be other mechanisms for convergent melt flow in addition to the model shown here (see, for example, refs 23, 24) these results probably rule out significant convergent solid flow within the partially molten region.

The results we report here are consistent with both geochemical and geophysical reasoning. Results from the seismic MELT experiment at 17° S on the southern EPR suggest that melt is distributed over a wide area at depth⁸. The simultaneous existence of a narrow neovolcanic zone at the surface implies some mechanism for lateral melt migration in the MELT area. Although seismic or electromagnetic techniques can sense the presence of melt, they probably cannot distinguish the relative geometries of melt and solid flow fields. However, our results suggest that geochemistry can make this distinction, because the geometry of melt and solid flow controls the mixing of melts and their interaction with different solid residues. We interpret the geochemical signatures at 12° N as the most direct evidence yet for lateral melt migration. □

Received 19 July; accepted 27 September 1999.

1. Spiegelman, M. & McKenzie, D. Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth Planet. Sci. Lett.* **83**, 137–152 (1987).
2. Phipps Morgan, J. Melt migration beneath mid-ocean spreading centers. *Geophys. Res. Lett.* **14**, 1238–1241 (1987).
3. Scott, D. R. & Stevenson, D. J. A self-consistent model of melting, magma migration and buoyancy-driven circulation beneath mid-ocean ridges. *J. Geophys. Res.* **94**, 2973–2988 (1989).
4. Buck, W. R. & Su, W. Focused mantle upwelling below mid-ocean ridges due to feedback between viscosity and melting. *Geophys. Res. Lett.* **16**, 641–644 (1989).
5. Scott, D. R. in *Mantle Flow and Melt Generation at Mid-ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 327–352 (Geophysical Monogr. Vol. 71, American Geophysical Union, Washington DC, 1992).
6. Cordery, M. J. & Phipps Morgan, J. Convection and melting at mid-ocean ridges. *J. Geophys. Res.* **98**, 19477–19503 (1993).
7. Sparks, D. W. & Parmentier, E. M. The structure of three-dimensional convection beneath oceanic spreading centres. *Geophys. J. Int.* **112**, 81–91 (1993).
8. Forsyth, D. W. *et al.* Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science* **280**, 1215–1217 (1998).
9. Spiegelman, M. Geochemical consequences of melt transport in 2-D: The sensitivity of trace elements to mantle dynamics. *Earth Planet. Sci. Lett.* **139**, 115–132 (1996).
10. Reynolds, J. R., Langmuir, C. H., Bender, J. F., Kastens, K. A. & Ryan, W. B. F. Spatial and temporal variability in the geochemistry of basalts from the East Pacific Rise. *Nature* **359**, 493–499 (1992).
11. Reynolds, J. R. *Segment Scale Systematics of Mid-ocean Ridge Magmatism and Geochemistry*. (Thesis, Columbia Univ./LDEO, (1995).
12. Reynolds, J. R. & Langmuir, C. H. Identification and implications of off-axis lava flows around the East Pacific Rise. *Earth Planet. Sci. Lett.* (in the press).
13. Plank, T. & Langmuir, C. H. Effects of the melting regime on the composition of the oceanic crust. *J. Geophys. Res.* **97**, 19749–19770 (1992).
14. McKenzie, D. & O'Nions, R. K. Partial melt distributions from inversion of rare earth element concentrations. *J. Petrol.* **32**, 1021–1091 (1991).
15. Monti, S., Gente, P. & Maze, J. P. *Carte Bathymétrique de la Dorsale Est-Pacifique de 10° Nord à 14° 30' Nord* (IFREMER, Brest, 1987).
16. Crane, K. Structural evolution of the East Pacific Rise axis from 13° 10' N to 10° 35' N; interpretation from Sea-MARC I data. *Tectonophysics* **136**, 65–124 (1987).
17. Perfit, M. R. & Chadwick, W. W. Jr in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, R. W., Delaney, P. T., Karson, J. A. & Lagabriele, Y.) 59–115 (Geophysical Monogr. Ser., Vol. 106, American Geophysical Union, Washington DC, 1998).
18. Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
19. Perfit, M. R. *et al.* Small-scale spatial and temporal variations in midocean ridge crest magmatic processes. *Geology* **22**, 375–397 (1994).
20. Perfit, M. R. *et al.* Detailed petrological and geochemical studies of on- and off-axis lavas from the East Pacific Rise; 9° 30'–10° N. *Eos* **75**, 601 (1994).
21. Weaver, J. S. & Langmuir, C. H. Calculation of phase equilibrium in mineral-melt systems. *Comput. Geosci.* **16**, 1–19 (1990).
22. Goldstein, S. J., Perfit, M. R., Batiza, R., Fornari, D. J. & Murell, M. T. Off-axis volcanism at the East Pacific Rise detected by uranium-series dating of basalts. *Nature* **367**, 157–159 (1994).
23. Sparks, D. W. & Parmentier, E. M. Melt extraction from the mantle beneath spreading centers. *Earth Planet. Sci. Lett.* **105**, 368–377 (1991).
24. Spiegelman, M. Physics of melt extraction: Theory, implications and applications. *Phil. Trans. R. Soc. Lond. A* **342**, 23–41 (1993).

Acknowledgements

We thank M. Perfit for comments and suggestions, and C. H. Langmuir and P. Asimow for discussions. This work was supported by NSF-OCE.

Correspondence and requests for materials should be addressed to M.S. (e-mail: mspieg@ldeo.columbia.edu).

Energetic constraints on the diet of terrestrial carnivores

Chris Carbone^{*†}, Georgina M. Mace^{*}, S. Craig Roberts^{*} & David W. Macdonald[†]

^{*} Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

[†] Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Species in the mammalian order Carnivora exhibit a huge diversity of life histories with body sizes spanning more than three orders of magnitude. Despite this diversity, most terrestrial carnivores can be classified as either feeding on invertebrates and small vertebrates or on large vertebrates. Small carnivores feed predominantly on invertebrates probably because they are a superabundant resource (sometimes 90% of animal biomass¹⁻³); however, intake rates of invertebrate feeders are low, about one tenth of those of vertebrate feeders^{4,5}. Although small carnivores can subsist on this diet because of low absolute energy requirements, invertebrate feeding appears to be unsustainable for larger carnivores. Here we show, by reviewing the most common live prey in carnivore diets, that there is a striking transition from feeding on small prey (less than half of predator mass) to large prey (near predator mass), occurring at predator masses of 21.5–25 kg. We test the hypothesis that this dichotomy is the consequence of mass-related energetic requirements and we determine the predicted maximum mass that an invertebrate diet can sustain. Using a simple energetic model and known invertebrate intake rates, we predict a maximum sustainable mass of 21.5 kg, which matches the point where predators shift from small to large prey.

After allowing for shared evolutionary history, we found an overall positive correlation between carnivore body mass and the mass of their most common prey (comparative analysis by independent contrasts (CAIC)^{6,7}: prey mass = $1.19 \times$ predator mass; $n = 112$, $r = 0.45$; $P < 0.001$; Fig. 1). There was also a pronounced shift to larger prey items by carnivores at intermediate masses. We defined this shift point statistically as the point where the data can be split into two groups so as to minimize the sum of the mean squares. For this analysis, we excluded the sloth bear (*Ursus ursinus*),

which is clear outlier (see below; Fig. 1a), and found a shift in the data between 21.5 and 25 kg.

We conducted an analysis to predict the maximum body mass that can be sustained on invertebrate prey. Our analysis combined a net-rate model⁸ and estimates for mass-related metabolic rates⁹ and travel speeds¹⁰⁻¹², along with estimates of invertebrate intake rates (Table 1). We selected species that have a range of body sizes, habitats and feeding strategies, including specialist invertebrate feeders (badger¹³, *Meles meles*; bat-eared fox^{12,14}, *Otocyon megalotis*; aardwolf^{11,15} *Proteles cristatus*), a largely invertebrate feeder (banded mongoose¹⁶, *Mungos mungo*) and a generalist (golden jackal¹⁷, *Canis aureus*). We then compared the model's predictions of the maximum body masses that could be sustained on an invertebrate diet with the statistically defined shift point determined above.

The predictions of maximum sustainable mass from the net-rate model varied among species: 5.1 kg for bat-eared fox, 8.5 kg for banded mongoose, 13.0 kg for golden jackal, 20.9 kg for aardwolf, and 21.5 kg for badger (Fig. 2). The differences in species estimates is partly due to differences in feeding ecologies but may also result from practical difficulties in accurately measuring intake rates of very small prey. The highest values estimated from the aardwolf and badger are very similar and correspond closely with the mass at which carnivores shift to feeding on large prey. However, because some species may not be able to achieve such high intake rates, we have included both the minimum and maximum estimate in Fig. 1a–e. Although these extremes cover a range of 16.4 kg, this range is small compared with the total range in carnivore mass, and more than 3/4 of the species in the analysis fall outside of this range.

We summarized the most common prey masses of carnivores in relation to this maximum-mass estimate. Carnivores weighing 21.5 kg or less feed mostly on prey that is 45% or less of their own mass, whereas carnivores above this feed mostly on prey that is greater than 45% of their own mass (the percentage of carnivore mass was chosen to maximize the model fit). This classification holds for 92.1% of 139 species (92.9% of the smaller mass class, 82.6% of the larger mass class). We also summarized diet types of 158 species (including mainly vegetarian species and scavengers) according to this mass estimate. With all of the species combined, only about 1/4 of the species weighing 21.5 kg or less are purely vertebrate feeders, the rest are either omnivores (45%) (often with invertebrate prey included), purely invertebrate feeders (10%) or mixed invertebrate and small vertebrate feeders (19%). Just over half of the carnivores above this weight are purely vertebrate feeders (52%). The rest are omnivores, and bears account for most of these.

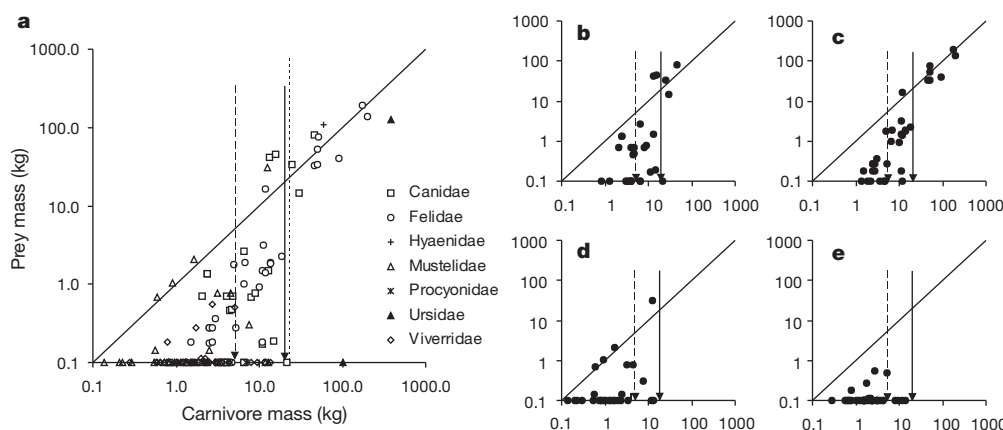


Figure 1 The mass of the most common prey plotted against carnivore mass. **a**, All families combined; **b**, Canidae; **c**, Felidae; **d**, Mustelidae; **e**, Viverridae. Average prey sizes of 0.1 kg or less were grouped into one prey-weight class. The diagonal line represents a slope of 1. The point that separates the prey masses into two groups while minimizing the sum of the mean squares is designated by the dotted line. The upper and

lower estimates of maximum carnivore mass (from Fig. 2) are designated by the vertical arrows. Above the 21.5-kg threshold, all species fit the model (except the sloth bear). Mustelidae and Viverridae do not exceed the threshold mass and almost all these species specialize on very small prey.

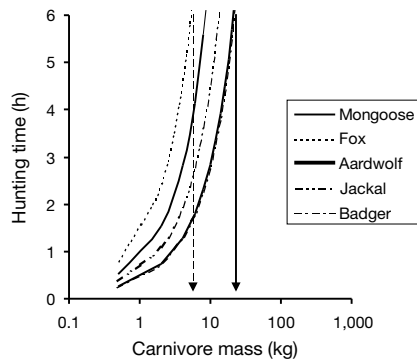


Figure 2 Predicted hunting time required to balance the net energy expenditure against carnivore body mass. As mass increases, the total energy expenditure increases, and so too must hunting time in order to balance the energy budget. Five predictions are presented based on the intake rate and travel speed of the banded mongoose, bat-eared fox, aardwolf, golden jackal and European badger (the aardwolf and badger cases overlap). The high (badger, solid line) and low (fox, dashed line), based on a hunting time of 6 h are designated by the vertical arrows.

Among canids and felids, all species weighing over 21.5 kg are purely vertebrate feeders. Below this weight, 53% of the species are purely vertebrate feeders; the remaining are omnivores (29%), purely invertebrate feeders (2%), and invertebrate and vertebrate feeders (16%).

To see how the variation in model parameters affected our predictions, we calculated the predicted maximum sustainable mass for species in Table 1, over a full range of foraging times (Fig. 3). The overall maximum mass predicted was 52 kg (equivalent to a badger feeding 24 h day⁻¹). This suggests that invertebrates cannot sustain the very largest carnivores even under exceptional prey densities and foraging times. In addition, when estimates of travel speeds vary by $\pm 50\%$, an average variation occurs in the predicted maximum masses of only -12.3% and $+16.8\%$.

We propose that the maximum size that can be sustained on an invertebrate diet is roughly 21.5 kg, and that this results from the simple energetic constraint of relying on a small particle-sized food source. This can explain the general absence of medium to large carnivores (more than 21.5 kg) feeding on very small prey (Fig. 1a). Although there are conditions under which higher maximum body mass might be predicted, these rely on exceptionally high intake rates that are patchily or temporally limited and not relevant to long-term energetic constraints. For example, badger intake rates may greatly exceed the long-term average used in our analysis¹³, but this only occurs seasonally and in areas with intensive farming^{13,18}. There are also exceptions to the overall pattern that we describe

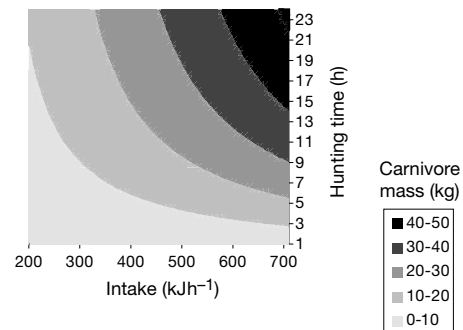


Figure 3 A contour plot of the maximum carnivore body mass predicted from the net-rate model for a range of invertebrate intake rates and hunting times. The range of intake rates used is based on the examples in Table 1.

here. Aquatic carnivores were excluded because of different thermoregulatory constraints and lower transport costs due to buoyancy¹⁹. Bears are also generally an exception; only the polar bear (*Ursus maritimus*) and the sloth bear feed predominantly on live animal foods. The polar bear is consistent with our model predictions, being large and feeding on large prey. The sloth bear probably maintains its large size both through its unusual ability to suck termites from the core of excavated colonies and by feeding heavily on fruits when available²⁰. In conclusion, our results provide a powerful framework for interpreting general trends in diet choice and evolution of body size in this diverse group. Our results also support the idea that large carnivores are a distinct group, both ecologically and physiologically, rather than simply a scaled-up version of a small carnivore, and that energetic constraints may be important for understanding their vulnerability²¹. □

Methods

Carnivore diet

To examine dietary preferences, we collated data on the diets and body sizes of 158 species of terrestrial carnivores from the literature (the main sources were refs 22–27). We excluded aquatic carnivores, including pinnipeds and other species that are largely dependent on aquatic habitat and prey (for example, otters, mink, and three species of viverrids, see ref. 19). We recorded the body mass (mean mass, or midpoint of stated range) of each species and that of its common prey. We conducted a separate analysis on 139 species excluding those which feed primarily on carrion or vegetable foods. Where there was no information on prey preference, mean prey size was calculated from the prey species listed. For small prey, where species weights were not given, masses were estimated from average weights for the genus, or were estimated at 0.001 kg for insects, 0.005 kg for aquatic invertebrates, 0.1 kg small rodents and birds and 0.5 kg for small mammals.

Invertebrate intake rates

Estimates of energetic intake rates on a diet of invertebrates were based on studies of the banded mongoose¹⁶, the bat-eared fox^{13,15}, the aardwolf^{11,15} the golden jackal¹⁷ and the

Table 1 Values used in the estimates of energy expenditure while hunting (E_h) and intake rates (I) for five species of carnivores

Species	Species mass (kg)	Speed* (m s ⁻¹)	Prey mass (g)	Estimated intake rate† (I) (kJ h ⁻¹)	Notes
Banded mongoose	1.5	0.276	529 Insects spp (0.0085) 102 Diplopoda (0.75) 5 Arachnida (1.0) 1 Mollusca (3.25)	347.0	Invertebrates captured over 590 min
Bat-eared fox	4.1	0.348	Termites 82% (0.0075) Beetles 12.3% (0.17)	234.0	% stomach volume Capture rate of 12 min
Aardwolf	10.0	0.434	Termites (<i>Trinervitermes</i> and <i>Hodotermes</i> spp.)	693.4	Average of winter and summer rates
Golden jackal	11.0	0.454	Beetles (0.17)	480.0	1.23 min
European badger	13.0	0.463	Worms (10,460 J each)	709.2	1.13 min (long-term average)

All travel speed were calculated from the equation of travel speed in relation to body mass (see Methods).

* These values are close to observed speeds (for example, mongoose (D. De Luca, unpublished data), bat-eared fox^{12,14} and aardwolf^{11,19}).

† Energetic values for the invertebrates were estimated at 38,258 J g⁻¹ (ref. 4), except for earthworms, for which a per-item energetic value is given¹⁸ and termites (ref. 15). Invertebrate weights were obtained from published references; and D. De Luca and P. Pearce-Kelly, personal communication.

European badger¹³ (Table 1). The estimates of intake rates and travel speeds used in the calculations are presented in Table 1.

Net-rate model

We used the net-rate model assuming a balanced energy expenditure and solving for time hunting, T_h (in h) (see ref. 8), $T_h = 24E_r/(I + E_r - E_h)$, where E_h and E_r are the energetic expenditure rates (in kJ h^{-1}) while hunting and resting and I is the intake rate while hunting (in kJ h^{-1}). The resting and hunting rates of energy expenditure (in Watts) for a carnivore of a given mass (M_b) (in kg) was calculated from the equations, $E_r = 3.34 M_b^{0.73}$; and $E_h = 10.7 M_b^{0.684} \nu + 6.03 M_b^{0.697}$ where ν is the average speed (in m s^{-1}) during hunting⁹. We estimate ν using the equation $V = 0.25 M_b^{0.24}$ (ref. 10). The constant, representing m s^{-1} (for a 1 kg animal), is based on travels speeds in the aardwolf¹, but this equation produces travel speeds within observed limits for other carnivore species listed in Table 1 (refs 11, 12; D. De Luca, unpublished data). Our calculations are based on a foraging time of 6 h which represents typical foraging times for a number of invertebrate feeding carnivores (for example, see refs 4, 11, 28, 29). Additional estimates of maximum body mass were made for varying intake rates and a full range of hunting times (Fig. 3).

Received 30 June; accepted 6 September 1999.

- Wilson, E. O. The little things that run the world: the importance of conservation of invertebrates. *Conserv. Biol.* **1**, 344–346 (1987).
- Wilson, E. O. *Success and Dominance in Ecosystems: the Case of Social Insects*. (Ecology Institute, Olendorf, Luhe, Germany, 1990).
- May, R. M. How many species are there on earth? *Science* **241**, 1441–1449 (1988).
- Waser, P. M. Small nocturnal carnivores: ecological studies in the Serengeti. *Afr. J. Ecol.* **18**, 167–185 (1980).
- Schaller, G. B. *The Serengeti Lion: a Study of Predator–Prey Relations* 1–480 (Univ. Chicago Press, Chicago, 1972).
- Purvis, A. & Rambaut, A. Comparative analysis by independent contrasts: an Apple Macintosh application for analysing comparative data. *Comput. Appl. Biosci.* **11**, 247–251 (1995).
- Bininda-Emonds, O. R. P., Gittleman, J. L. & Purvis, A. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol. Rev.* **74**, 143–175 (1994).
- Gorman, M. L., Mills, M. G., Raath, J. P. & Speakman, J. R. High hunting costs make African wild dogs vulnerable to kleptoparasitism by hyaenas. *Nature* **391**, 479–481 (1998).
- Schmidt-Nielsen, S. *Scaling: Why Is Animal Size So Important?* 1–241 (Cambridge Univ. Press, Cambridge, 1984).
- Taylor, C. R., Heglund, N. C. & Maloiy, G. M. O. Energetics and mechanics of terrestrial locomotion. I. Metabolic energy consumption as a function of speed and body size in birds and mammals. *J. Exp. Biol.* **97**, 1–21 (1982).
- Richardson, P. R. K. Food-consumption and seasonal-variation in the diet of the Aardwolf *Proteles cristatus* in Southern-Africa. *Z. Säugetierk.* **52**, 307–325 (1987).
- Skinner, J. D. & Smithers, R. H. N. *Mammals of the Southern African Subregion* (Univ. Pretoria, South Africa, 1990).
- Kruuk, H. *The Social Badger* (Oxford Univ. Press, Oxford, 1989).
- Malcolm, J. R. Socio-ecology of bat-eared foxes, *Otocyon megalotis*. *J. Zool. Lond.* **208**, 743–744 (1986).
- Williams, J. B., Anderson, M. D., Richardson, P. R. K. Seasonal differences in field metabolism, water requirements, and foraging behavior of free-living aardwolves. *Ecology* **78**, 2588–2602 (1997).
- Rood, J. P. Population dynamics and food habits of banded mongoose. *E. Afr. Wildl. J.* **13**, 89–111 (1975).
- Lamprecht, J. On diet, foraging behaviour and interspecific food competition of jackals in the Serengeti National Park, East Africa. *Z. Säugetierk.* **43**, 210–223 (1978).
- Macdonald, D. W. The Red Fox, *Vulpes vulpes*, as a predator upon earthworms, *Murchieus terrestris*. *Z. Tierpsychologie* **52**, 171–200 (1980).
- Bininda-Emonds, O. R. P. Towards larger phylogenies: examples within the Carnivora (Mammalia). Thesis, Univ. Oxford (1998).
- Joshi, A. R., Garshelis, D. L. & Smith, J. L. D. Seasonal and habitat-related diets of sloth bears in Nepal. *J. Mammology* **78**, 584–597 (1997).
- Purvis, A., Mace, G. M. & Gittleman, J. L. in *Carnivore Conservation* (eds Gittleman, J. L., Funk, S., Macdonald, D. W. & Wayne, R. K.) (Cambridge Univ. Press, Cambridge, in the press).
- Gittleman, J. L. in *Carnivore Behaviour, Ecology and Evolution* (ed. Gittleman, J. L.) 183–207 (Chapman Hall, London, 1989).
- Gittleman, J. L. Carnivore body size: Ecological and taxonomic correlates. *Oecologia* **67**, 540–554 (1995).
- Nowak, R. M. *Walker's Mammals of the World* (The Johns Hopkins Univ. Press, Baltimore, 1991).
- Estes, R. D. *The Behaviour Guide to African Mammals: Including Hoofed Mammals, Carnivores, Primates* 1–611 (Univ. California Press, Berkeley and Los Angeles, 1991).
- Macdonald, D. W. *The Encyclopedia of Mammals* (Andromeda Oxford Ltd, Abingdon, UK, 1995).
- Kingdon, J. *The Kingdon Field Guide to African Mammals* 1–464 (Academic, London, 1997).
- Kavanau, J. L. & Ramos, J. Influences of light on activity and phasing of carnivores. *Am. Nat.* **109**, 391–418 (1975).
- Griffiths, D. Foraging costs and relative prey size. *Am. Nat.* **116**, 743–752 (1980).

Acknowledgements

We thank R. McNeill Alexander, D. Johnson, M. Rowcliffe, T. Coulson, J. Du Toit and D. De Luca for helpful discussions and comments on earlier drafts of the manuscript. G.M.M. Thanks the NERC.

Correspondence and requests for materials should be addressed to C.C. (e-mail: chris.carbone@ioz.ac.uk).

Total lipid energy as a proxy for total egg production by fish stocks

C. Tara Marshall*, Nathalia A. Yaragina†, Yvan Lambert‡ & Olav S. Kjesbu*

* Institute of Marine Research, PO Box 1870 Nordnes, N-5817 Bergen, Norway

† Polar Research Institute of Marine Fisheries and Oceanography, 6 Knipovich Street, Murmansk, 183763, Russia

‡ Ministère des Pêches et des Océans, Institut Maurice-Lamontagne, C.P. 1000, Mont-Joli, G5H 3Z4, Canada

The indeterminate relationship between the total biomass of mature fish (spawner biomass) and the number of offspring produced (recruitment) has puzzled population dynamicists¹ and impeded fisheries management². The relationship assumes that spawner biomass (in tonnes) is proportional to the total number of eggs produced (TEP) by the stock³, an assumption under increasing challenge^{4–8}. Most stocks require proxies for TEP because contemporary and/or historical fecundity data are lacking. Here we show a positive association between recruitment and the liver weights of spawners in the Barents Sea cod stock which suggests that recruitment is constrained by the amount of lipid energy stored in the liver. This stimulated our interest in estimating total lipid energy (TLE; in kilojoules) for mature females in the stock. We examined the suitability of TLE as a proxy through correlation and simulation analyses. The results indicate that TLE is proportional to TEP and exhibits a similar response to varying food abundance. Replacing spawner biomass with more accurate measures of reproductive potential is essential to developing a rational basis for stock conservation⁹. Correctly specifying the first-order maternal effect on TEP is a prerequisite to detecting environmental and ecological effects on recruitment¹⁰.

The protein and lipid reserves used for metabolism, gonad development and spawning behaviour of fish co-vary in response to the abundance of food¹¹. In gadoid species, such as Atlantic cod, lipids are stored primarily in the liver¹² making liver weight a rapid and inexpensive measure of spawner quality. Russia began recording liver weights of Barents Sea cod in 1927. Since 1967, liver weights have been recorded monthly for fish in 10-cm length classes¹³. These data are archived as the liver condition index (LCI) which expresses

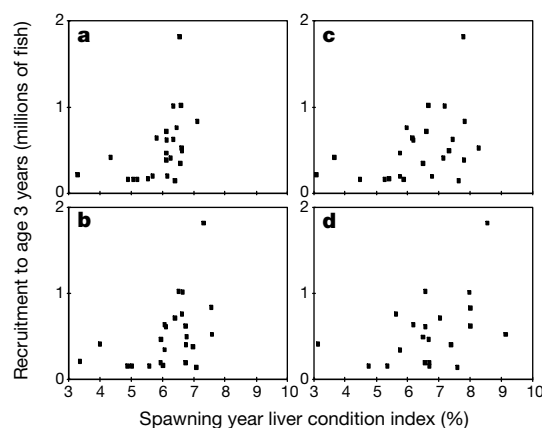


Figure 1 Empirical relationship between recruitment to age three years¹⁸ and the mean liver condition index estimated for a single spawning year (defined as the mean of monthly values recorded for July the preceding year through to June of the spawning year). The spawning years 1968–1994 are represented. Length classes of fish: **a**, 51–60 cm; **b**, 61–70 cm; **c**, 71–80 cm; **d**, 81–90 cm.

liver weight as a percentage of the total body weight. The annual mean LCI is positively correlated with the abundance of capelin¹³, a lipid-rich prey species common in arcto-boreal ecosystems. The liver energy content (LEC; in kilojoules per gram) can be determined from the LCI using the non-linear regression model ($r^2 = 0.64$, $n = 407$, $P < 0.001$):

$$\text{LEC} = 24.77(1 - \exp^{-0.52(\text{LCI} - 0.48)}) \quad (1)$$

Equation 1 was developed for Gulf of St Lawrence cod having LCI values of approximately 1–8% (ref. 12). Here we assume that it applies to Barents Sea cod. This asymptotic equation indicates that the increase in LEC is negligible beyond an LCI value of approximately 5% while below that value the LEC decreases rapidly.

Food-dependent variation in stored lipid energy influences the reproductive potential of individual fish^{14–16}. We have found that at the stock level it may also constrain recruitment. Poor recruitment to the Barents Sea cod stock occurs when the LCI is less than 6% (Fig. 1), a value which is close to the level when the LEC begins to decline rapidly (equation 1). This empirical result is consistent with a positive correlation between recruitment of South African anchovy and the quantity of oil contained in the commercial anchovy catch expressed as a percentage of meal production¹⁷.

Positive associations between recruitment and lipid-based indices suggest that TEP may be linked to stored lipid energy at the stock level. Both the LCI and the oil/meal ratio indicate the quality of spawners but not their quantity. A suitable proxy for TEP should incorporate both aspects of reproductive potential in proportion to their degree of influence on TEP. Accordingly, we used survey data describing abundance and demographics of the Barents Sea cod stock⁶ to estimate the TLE for 1985–1996 (1997 and 1998 were excluded because of restricted spatial coverage in those years) as follows. For each survey year the biomass of mature females in 10-cm length classes (50–140 cm) was estimated as the product of length-specific estimates of the number of cod (n_1), proportion of females (s_1), proportion of females that are mature (m_1), and body weight (w_1 ; in grams). This value was multiplied by the annual mean LCI for that length class (LCI_1 ; expressed as a proportion of total body weight, with missing values being approximated using either interpolated values or the nearest observed values) to give the liver biomass of mature females in each length class. The TLE was calculated by integrating the product of liver biomass and length-specific LEC (LEC_1 ; predicted from LCI_1 by equation (1)) across length.

We find a highly significant, linear relationship ($r^2 = 0.97$,

$n = 21$, $P < 0.001$) between TEP and TLE, with an intercept that is not significantly different from zero ($P > 0.05$):

$$\text{TEP} = 407\text{TLE} \quad (2)$$

Thus, at the stock level 2.12 kJ of liver energy is proportional to 1000 eggs. By exhibiting a linear relationship with TEP which passes through the origin, TLE meets the criteria for direct proportionality. Estimates of spawner biomass from standard analytical assessments¹⁸ for the same time period do not meet these criteria⁶.

We further evaluated the TLE using simulations which held the quantity of females constant and modelled m_1 , w_1 , length-specific fecundity (f_1), and LCI_1 as functions of capelin biomass (Table 1). Each simulation estimated three output variables (TEP, TLE and spawner biomass) for seven capelin biomasses spanning the range observed in the Barents Sea¹⁹. Twelve separate simulations were run using observed n_1 values for 1985–1996. To remove an unimportant source of variation each simulation used the mean of s_1 values for 1985–1996 ($\bar{s}_1$) to partition n_1 by sex. The TEP was calculated by integrating the product of n_1 , $\bar{s}_1$, m_1 , and f_1 across length. The TLE was calculated by integrating the product of n_1 , $\bar{s}_1$, m_1 , w_1 , LCI_1 , and LEC_1 across length. Spawner biomass was calculated by integrating the product of n_1 , m_1 and w_1 across length. Thus, spawner biomass makes no distinction between males and females, which is consistent with the standard analytical assessment for this stock¹⁸. Excluding m_1 , which is common to all three output variables, TEP and spawner biomass have one food-dependent term each (f_1 and w_1 , respectively), while TLE has three (w_1 , LCI_1 , and LEC_1 predicted from LCI_1). The food-dependent terms had either modelled or estimated errors (Table 1). For each simulation Monte Carlo sampling was used to estimate the mean and standard deviation (s.d.) for each output variable at each capelin biomass. Convergence (less than 1.5% change in both the mean and the s.d.) was achieved within 1,500 iterations and 95% confidence intervals were approximated as the mean ± 2 s.d. To compare the magnitude of variation in TEP, TLE and spawner biomass on a common scale, the means were divided by the mean calculated for that simulation at the minimum capelin biomass.

In all twelve simulations the mean values of TEP and TLE at high capelin biomasses are significantly higher than those at the minimum capelin biomass (100,000 t). The differences become significant at capelin biomasses of either 10^6 t or 4×10^6 t for TEP and at a capelin biomass of 10^6 t for TLE. By either of these measures the reproductive potential of a fixed quantity of mature females is significantly higher when food is abundant. Spawner biomass at the maximum capelin biomass is not significantly different from spawner biomass at the minimum capelin biomass in three of the

Table 1 Models used to estimate input variables in the simulation analysis

Input variable	Calibration period	Model type	Independent variables	Standard deviation	r^2	P
m_1	1985–1996 (ref. 6)	Logistic	Intercept Length	*	NA	NA NA NA
w_1	1985–1996 (ref. 6)	Multiple regression	Capelin biomass Intercept ln (length)	0.17	0.93	<0.0001 <0.0001 <0.0001
f_1	1986–1991, except 1990 (ref. 29)	Multiple regression	ln (capelin biomass) Interaction Intercept ln (length)	0.32	0.90	<0.0001 <0.0001 <0.0001 <0.0001
LCI_1	1973–1996 (ref. 13)	ANCOVA	ln (capelin biomass) Interaction Intercept Length ln (capelin biomass)	0.67	0.66	0.0003 0.004 0.017 <0.0001

Length refers to the mid-point of 5-cm length classes in the range 50–140 cm, except LCI_1 , which is resolved by 10-cm length classes. Estimates of capelin biomass were obtained from annual acoustic surveys¹⁹. Values of LCI_1 used to calibrate the model correspond to the mean LCI estimated for a single calendar year. Variables which were transformed to natural logarithms are identified as ln(variable name). In the ANCOVA model, length classes and ln(capelin biomass) were the treatment variable and covariate, respectively. Errors were normally distributed with a mean of 0 and the standard deviations described above. NA indicates unavailable data. r^2 , coefficient of correlation (squared); P , significance of probability.

* Standard deviation (s.d.) was set to a maximum value at the length at which $m_1 = 0.5$ and decreased linearly to 0 both below that length (s.d. = 0 when length was 50 cm) and above that length (s.d. = 0 when length ≥ 100 cm).

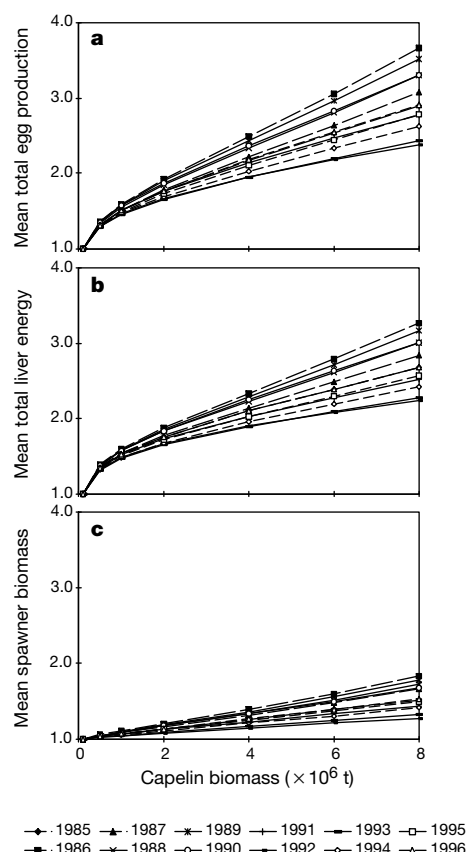


Figure 2 Simulated relationship between mean values of the three output variables expressed relative to the value at the minimum capelin biomass. **a**, Total egg production; **b**, total liver energy; **c**, spawner biomass. Values were estimated using the length compositions observed in surveys of the Barents Sea and Lofoten region for 1985–1996 (ref. 6).

twelve simulations (1991–1993). Otherwise, differences become significant at capelin biomasses of either 4×10^6 or 6×10^6 t.

For a fixed quantity of mature females, TEP is 2.4–3.7 times higher when capelin biomass is at a maximum, TLE is 2.2–3.3 times higher, and spawner biomass is 1.3–1.8 times higher (Fig. 2). The relative increase in TLE, driven by food-dependent increases in w_1 , LCI_1 and LEC_1 , closely tracks the increase in TEP. Both indicate a two- to four-fold increase in the reproductive potential of a fixed quantity of spawners as food increases to maximum observed values. The relative increase in spawner biomass, driven by w_1 alone, is approximately half that observed in TEP. Thus, spawner biomass lacks the dynamic range of TEP. The inherent insensitivity of spawner biomass to changes in reproductive potential will be exacerbated by using constant values of length- or age-specific weight and/or maturity. Such is the case for 36 of the 52 annual estimates of spawner biomass for Barents Sea cod¹⁸.

Given these intrinsic and imposed constraints on the dynamics of spawner biomass it is not surprising that it performs poorly both as a proxy for TEP (ref. 6) and as a predictor of recruitment^{1,2}. The TEP is a direct measure of reproductive potential which shows promise for predicting recruitment^{6,20,21}. However, TLE is a practical alternative when the fecundity data required to estimate TEP are unavailable and a time series quantifying stored lipid energy at critical points in the annual spawning cycle can be developed at minimal cost. Data collected for industrial purposes (such as LCI and oil/meal ratios) can be used to reconstruct historical variation in reproductive potential. Conceptually, TLE is appealing because lipids represent a common “currency”¹⁶ linking food quality and

quantity to many aspects of fish nutrition²², reproduction^{14,15,23} and population dynamics²⁴. Analogues of TLE could be developed for use in ecosystem models tracing lipid flow among trophic levels^{25,26}. Given the importance of lipids in vertebrate²⁷ and invertebrate²⁸ reproduction, our use of stored lipid energy to predict features of population dynamics seems to be a widely applicable approach. □

Received 15 July; accepted 15 September 1999.

- Rothschild, B. J. *Dynamics of Marine Fish Populations* (Harvard Univ. Press, Cambridge, 1986).
- Hilborn, R. & Walters, C. J. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty* (Chapman and Hall, New York, 1992).
- Beverton, R. J. H. & Holt, S. J. *On the Dynamics of Exploited Fish Populations* (Fish. Invest. Ministry of Agriculture Fish. Food Ser. II, Salmon Freshwater Fisheries, London, 1957).
- Trippel, E. A., Kjesbu, O. S. & Solemdal, P. in *Early Life History and Recruitment in Fish Populations* (eds Chambers, R. C. & Trippel, E. A.) 31–62 (Chapman and Hall, London, 1997).
- Marteinsdottir, G. & Thorarinnsson, K. Improving the stock–recruitment relationship in Icelandic cod (*Gadus morhua* L.) by including age diversity of spawners. *Can. J. Fish. Aquat. Sci.* **55**, 1372–1377 (1998).
- Marshall, C. T., Kjesbu, O. S., Yaragina, N. A., Solemdal, P. & Ulltang, Ø. Is spawner biomass a sensitive measure of the reproductive and recruitment potential of Northeast Arctic cod? *Can. J. Fish. Aquat. Sci.* **55**, 1766–1783 (1998).
- Lowerre-Barbieri, S. K., Lowerre, J. M. & Barbieri, L. R. Multiple spawning and the dynamics of fish populations: inferences from an individual-based simulation model. *Can. J. Fish. Aquat. Sci.* **55**, 2244–2254 (1998).
- Marshall, C. T. & Frank, K. T. The effect of interannual variation in growth and condition on haddock recruitment. *Can. J. Fish. Aquat. Sci.* **56**, 347–355 (1999).
- Painting, S. J. *et al.* Environmental and biological monitoring for forecasting anchovy recruitment in the southern Benguela upwelling region. *Fisheries Oceanography* **7**, 364–374 (1998).
- Jarre-Teichmann, A. *et al.* Stock-recruitment relationships for cod (*Gadus morhua callarias* L.) in the central Baltic Sea incorporating environmental variability. *ICES J. Mar. Sci.* (in the press).
- Lambert, Y. & Dutil, J.-D. Condition and energy reserves of Atlantic cod (*Gadus morhua*) during the collapse of the northern Gulf of St. Lawrence stock. *Can. J. Fish. Aquat. Sci.* **54**, 2388–2400 (1997).
- Lambert, Y. & Dutil, J.-D. Can simple condition indices be used to monitor and quantify seasonal changes in the energy reserves of Atlantic cod (*Gadus morhua*)? *Can. J. Fish. Aquat. Sci.* **54** (Suppl. 1), 104–112 (1997).
- Yaragina, N. A. & Marshall, C. T. Trophic influences on seasonal and interannual variation in the liver condition index of Northeast Arctic cod (*Gadus morhua*). *ICES J. Mar. Sci.* (in the press).
- Henderson, B. A., Wong, J. L. & Nepszky, S. J. Reproduction of walleye in Lake Erie: allocation of energy. *Can. J. Fish. Aquat. Sci.* **53**, 127–133 (1996).
- Henderson, B. A. & Wong, J. L. Control of lake trout reproduction: role of lipids. *J. Fish Biol.* **52**, 1078–1082 (1998).
- Adams, S. M. in *Lipids in Freshwater Ecosystems* (eds Arts, M. T. & Wainman, B. C.) 132–160 (Springer, New York, 1999).
- Boyd, A. J., Shannon, L. J., Schülein, F. H. & Taunton-Clark, J. in *Global versus Local Changes in Upwelling Systems* (eds Durand, M.-H. *et al.*) 195–209 (Éditions de l'Orstom, Paris, 1998).
- Report of the Arctic Fisheries Working Group ICES CM 1999/ACFM:3 (International Council for the Exploration of the Sea, Copenhagen, 1999).
- Gjoseter, H. The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. *Sarsia* **83**, 453–496 (1998).
- Huggett, J. A., Boyd, A. J., Hutchings, L. & Kemp, A. D. Weekly variability of clupeoid eggs and larvae in the Benguela jet current: implications for recruitment. *S. Afr. J. Mar. Sci.* **19**, 197–210 (1998).
- Mukhina, N. V. The use of Russian ichthyoplankton survey data in the forecasting of recruitment to Arcto-Norwegian cod stock. *ICES CM 1999/Y:15* (International Council for the Exploration of the Sea, Copenhagen, 1999).
- Sargent, J. R., Bell, J. G., Bell, M. V., Henderson, R. J. & Tocher, D. R. Requirement criteria for essential fatty acids. *J. Appl. Ichthyol.* **11**, 183–198 (1995).
- Pickova, J., Dutta, P. C., Larsson, P.-O. & Kiessling, A. Early embryonic cleavage pattern, hatching success, and egg-lipid fatty acid composition: comparison between two cod (*Gadus morhua*) stocks. *Can. J. Fish. Aquat. Sci.* **54**, 2410–2416 (1997).
- Bell, M. V. & Sargent, J. R. Lipid nutrition and fish recruitment. *Mar. Ecol. Prog. Ser.* **134**, 315–316 (1996).
- Lawson, J. W., Magalhães, A. M. & Miller, E. H. Important prey species of marine vertebrate predators in the northwest Atlantic: proximate composition and energy density. *Mar. Ecol. Prog. Ser.* **164**, 13–20 (1998).
- Olsen, Y. in *Lipids in Freshwater Ecosystems* (eds Arts, M. T. & Wainman, B. C.) 161–202 (Springer, New York, 1999).
- Frisch, R. E. Fatness and fertility. *Sci. Am.* **258**, 88–95 (1988).
- Boersma, M. & Vijverberg, J. Resource depression in *Daphnia galeata*, *Daphnia cucullata* and their interspecific hybrid: life history consequences. *J. Plankton Res.* **16**, 1741–1758 (1994).
- Kjesbu, O. S., Withames, O. R., Solemdal, P. & Greer Walker, M. Temporal variations in the fecundity of Arcto-Norwegian cod (*Gadus morhua*) in response to natural changes in food and temperature. *J. Sea Res.* **40**, 303–321 (1998).

Acknowledgements

We thank the staff at the Polar Research Institute of Marine Fisheries and Oceanography who maintained the liver condition database. This work was supported by the Norwegian Research Council (CTM), the Norwegian Foreign Department (NAY), the Canadian Department of Fisheries and Oceans (YL) and the European Union (OSK).

Correspondence and requests for materials should be addressed to C.T.M. (e-mail: tara@imr.no).

Use of behavioural stochastic resonance by paddle fish for feeding

David F. Russell, Lon A. Wilkens & Frank Moss

Center for Neurodynamics, University of Missouri at St. Louis, St Louis, Missouri 63121, USA

Stochastic resonance is the phenomenon whereby the addition of an optimal level of noise to a weak information-carrying input to certain nonlinear systems can enhance the information content at their outputs^{1–4}. Computer analysis of spike trains has been needed to reveal stochastic resonance in the responses of sensory receptors^{5–7} except for one study on human psychophysics⁸. But is an animal aware of, and can it make use of, the enhanced sensory information from stochastic resonance? Here, we show that stochastic resonance enhances the normal feeding behaviour of paddlefish (*Polyodon spathula*)^{9,10}, which use passive electroreceptors^{11,12} to detect electrical signals from planktonic prey¹³. We demonstrate significant broadening of the spatial range for the detection of plankton when a noisy electric field of optimal amplitude is applied in the water. We also show that swarms of *Daphnia* plankton are a natural source of electrical noise. Our demonstration of stochastic resonance at the level of a vital animal behaviour, feeding, which has probably evolved for functional success, provides evidence that stochastic resonance in sensory nervous systems is an evolutionary adaptation¹⁴.

Paddlefish feed on zooplankton, especially *Daphnia*⁹, in North American rivers where muddy turbidity limits vision. Instead, paddlefish use an electrosensory antenna to locate plankton¹³. The long flattened 'rostrum' in front of the mouth (Fig. 1a) is covered with tens of thousands of passive electroreceptors^{11,15} similar to the ampullae of Lorenzini of sharks and rays^{16,17}. The electroreceptors respond best to low frequency (0.5–20 Hz) external fields¹³, as

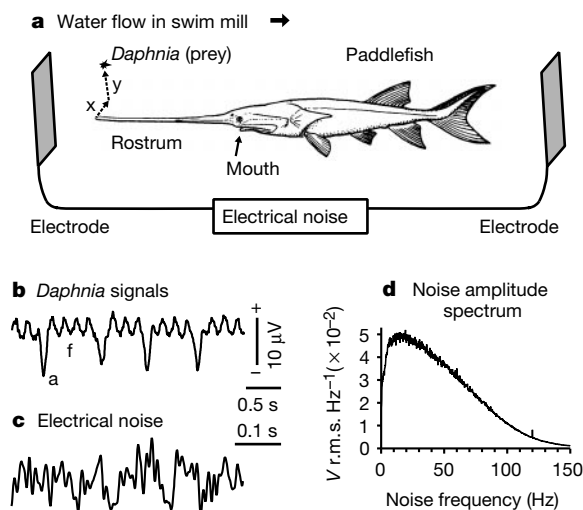


Figure 1 Experimental design. **a**, A small paddlefish (drawing reproduced with permission²⁶) fed on *Daphnia* plankton being swept towards it in a recirculating stream of water (swim mill), while noisy electrical current was passed through the water between large Ag–AgCl plate electrodes in front of and behind the fish, 95 cm apart. **b**, Electrical oscillations from a tethered *Daphnia* were correlated with rhythmic beating of the feeding legs (f) or antennae (a). The focal recording electrode was 1.2 mm behind the abdomen, with an agarose-coated plastic screen in between to block water currents. Bandwidth = 0–40 Hz. The reference electrode was distant, in 760 $\mu S \text{ cm}^{-1}$ water. **c**, **d**, Sample and amplitude spectrum of the electrical noise applied during feeding.

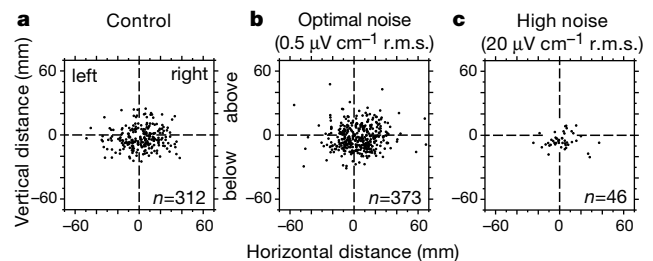


Figure 2 Spatial distributions of feeding strike locations at different noise levels. **a**, Control without noise; **b**, during an optimal amplitude of noise; **c**, during a high level of noise. Each point shows the location of an approaching *Daphnia* (plankton prey) which elicited a strike. The view is from the perspective of a fish 'looking' forward. The origin (0,0) corresponds to the midline-centre long axis of the rostrum. Distances from rostrum axis to *Daphnia* are relative. Data were pooled (see *n* values) from comparable-sized samples from each of four fish.

produced by *Daphnia* prey (Fig. 1b). We studied the role of electroreceptors in feeding behaviour by video observation of small paddlefish feeding on *Daphnia* in a recirculating stream of water (swim mill)^{10,13}. The spatial distribution of strike locations (Fig. 2a) surrounds the fish, but extends further horizontally than vertically. About 95% of *Daphnia* are taken at radial distances less than 40 mm. Plankton further from the rostrum are less likely to be detected, owing to the dipole-like electric field from *Daphnia*, which decreases approximately as the inverse cube of distance. Other sensory modalities, including vision and lateral line mechanosense, are not necessary for prey capture¹³.

Our hypothesis is that when a random electric field of optimal amplitude is applied to the environment in which a paddlefish is swimming and feeding, more distant plankton can be located and captured, compared with zero-field controls, owing to the dynamical process of stochastic resonance. Our hypothesis predicts that the spatial distribution of strike locations should become more spread out (show increased scatter or variance) when an optimal level of noise is presented. The only variable in our experiments was the selected amplitude (0.05–50 $\mu V \text{ cm}^{-1}$ r.m.s.) of a randomly varying electrical stimulus passed through the water in the swim mill, between plate electrodes in front of the fish and behind it (Fig. 1a). Noise of different amplitudes strongly affected the spatial distribution of strike locations. The scatter of the feeding strike locations was increased during presentation of 0.2–1 $\mu V \text{ cm}^{-1}$ noise, and was maximally increased during 0.5 $\mu V \text{ cm}^{-1}$ noise (Fig. 2b), compared with zero-field controls (Fig. 2a). This amplitude, 0.5 $\mu V \text{ cm}^{-1}$ r.m.s., was therefore defined to be the optimal noise amplitude. Higher levels of noise caused the distribution of strike locations to become compressed (show reduced scatter); that is, only nearby *Daphnia* elicited strikes (Fig. 2c).

Quantitative comparison of the spatial strike distributions during optimal noise or controls revealed that they differed chiefly in the vertical components of strike distances ($P_S(y)$, above or below the rostrum). A histogram of vertical strike distances (Fig. 3a) showed that with optimal noise (left side), the distribution became asymmetrical, the peak was shifted to ~ 10 mm below the rostrum, and there were additional outliers, compared to controls (right side). The mean and median shifted downwards only slightly (–0.92 and –1.37 mm, respectively). These distributions were normalized to the total number of strikes, so increased variance necessarily causes reduced amplitude near the centre. The increase in vertical spatial variance during optimal noise, compared to zero-field controls, was a robust finding in experiments with different fish over two years.

This increase is statistically significant ($P < 0.0001$ for equal dispersion, pooled data) as illustrated in Fig. 3b. We used the nonparametric Mood test for scale¹⁸, as the distributions (Fig. 3a)

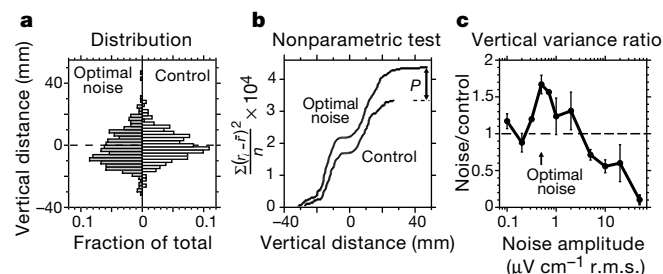


Figure 3 Spatial variance of strike locations. **a**, Normalized histograms of only the vertical components of fish-to-plankton strike distances (above or below the rostrum). **b**, Mood test¹⁸ showing increased scatter of vertical fish-to-plankton strike distances during optimal noise, seen as faster accumulation of rank variance than in control data. r_i , ranks of vertical distance values, after combining controls with optimal-noise data. $\bar{r}$, mean rank of combined data. n , sample size (312 for controls, 373 for optimal-noise data). P , probability was calculated for the difference between the final sums. **c**, Noise/control variance ratios (the conventional F statistic) for the vertical strike distances at different noise levels; mean $\pm$ s.e.m. for $n = 4$ fish, two runs per fish. Points without bars are from one fish.

were not gaussian. Three of the four fish reached significance when tested individually ($P < 0.05$, Mood; $P = 0.09$ for the fourth fish), from samples of comparable size (93 ± 27 s.d. values per fish during optimal noise, and 78 ± 28 in controls). Another nonparametric test¹⁹ set 95% confidence limits: the vertical spatial dispersion was 1.13–1.49 times larger during optimal noise than in controls (for pooled data, and asymmetrical distributions¹⁹). The effects of different noise amplitudes are portrayed in Fig. 3c using the conventional F ratio (variance of vertical strike distances during noise of a given amplitude, divided by the control variance, averaged over the four fish; a ratio of 1 would signify 'no effect'). By this measure, noise levels of $0.1 \mu\text{V cm}^{-1}$ or less had little effect. A maximum increase in the vertical spatial range, by ~ 1.6 -fold on average, or as much as 3-fold in some individual runs, was seen at 0.5 – $0.7 \mu\text{V cm}^{-1}$ noise. The F ratio declined to zero at higher noise levels (5 – $20 \mu\text{V cm}^{-1}$) as only nearby plankton were captured; this was attributable to masking of the electrical signals from *Daphnia*. The reduced dispersion at 5 – $20 \mu\text{V cm}^{-1}$ noise was significant by the Mood test ($P < 0.05$, pooled data). Noise of $50 \mu\text{V cm}^{-1}$ almost abolished feeding.

The regions in space around the fish where stochastic resonance was most effective were ascertained by replotted Figs 2a and b as normalized probability distributions (Fig. 4a and b), such that the volume under each surface sums to a total of one. Subtracting these surfaces (strike distribution with optimal noise minus the control distribution, Fig. 4c) revealed two bands of noise-enhanced strike probability, one above the rostrum and another larger band below the rostrum, around the mouth. This banding pattern reflects the enhancement of vertical strike distances.

The horizontal components of distance ($P_s(x)$, to the left or right of the rostrum) also showed increased variance during optimal noise, even though Mood statistical tests did not reach significance. For example, optimal noise caused the sharp central peak in controls (Fig. 4a) to be split into left and right shoulders (arrows, Fig. 4b), reflecting the increased horizontal variance of these normalized distributions. A plot of F ratios for the horizontal components at different noise amplitudes (not shown) was similar in form to Fig. 3c for vertical components, but reached a smaller maximum of 1.3 at the same optimal noise level.

Noise amplitudes of 0.1 – $2 \mu\text{V cm}^{-1}$, including optimal noise, did not significantly alter the number of *Daphnia* captured per minute, compared to controls, over the group of four fish studied ($P > 0.05$, analysis of variance (ANOVA); not illustrated). For two of the fish, $0.5 \mu\text{V cm}^{-1}$ noise did cause elevation of the capture rate above

controls by $\sim 50\%$ on average, or as much as twofold in some individual runs. However for the other two fish the capture rates at noise levels of 0.1 – $1 \mu\text{V cm}^{-1}$ were similar to control rates. For all the fish, the capture rate fell progressively to zero at higher noise amplitudes (5 – $50 \mu\text{V cm}^{-1}$).

If 'behavioural stochastic resonance' occurs in natural environments, there would need to be a source of electrical noise. Figure 5a–d shows that fluctuating electrical signals resembling noise are present near populations of *Daphnia*, which form dense swarms²⁰. The noise was detectable at least 5 cm from an enclosed swarm (Fig. 5d). Its amplitude declined approximately exponentially at distances beyond 2–3 cm (Fig. 5f), with a length constant of 2–3.2 cm. The noise amplitude was approximately gaussian, and increased with the population density as well as the body size of the individual *Daphnia*. Most components of the *Daphnia* noise (Fig. 5g) were within the bandwidth of paddlefish electroreceptors¹⁵. There are some differences between the *Daphnia* noise and the noise used in the experiment. For example, the same random signal was applied to all electroreceptors in the experiment, whereas for the *Daphnia* swarm the net signal is the sum of those produced by many spatially separate generators. A paddlefish swimming amongst a swarm of *Daphnia* will encounter such electrical noise. *Daphnia* may reveal their locations to paddlefish in two ways: individual *Daphnia* produce electrical signals (as in Fig. 1b), and populations produce background noise (as in Fig. 5) that could plausibly boost the sensitivity of electroreceptors, owing to stochastic resonance, during feeding in the wild. This may explain why paddlefish respond to external electrical noise: it is tied to their food source.

Three types of observation indicated that noise did not simply induce arousal^{21,22}. First, fish showed no visible response, such as special fin movements or startle flexures, when any of the noise amplitudes, even the highest, was switched on. Second, optimal noise did not affect the capture rate in two of the fish, even though these same fish did show increased spatial variance of strike locations. Third, the frequency of a different behaviour, in which fish would turn 180° (make a U-turn) and swim downstream, was also not significantly affected by the noise, compared to controls ($P > 0.05$, ANOVA; not illustrated). Another explanation for our data is that noise might cause the *Daphnia* to produce larger electrical signals. However, $50 \mu\text{V cm}^{-1}$ noise had no significant effect on the frequency or amplitude of the electrical signature arising from rhythmic beating motions of the feeding legs and antennae (recorded as in Fig. 1b), compared to controls without noise ($P > 0.05$, Kolmogorov–Smirnov test comparing power spectra).

We interpret these results as stochastic resonance increasing the sensitivity of peripheral^{3–7} electroreceptors, or acting within the brain^{12,23,24}, or both. Increased sensitivity allows the fish to detect

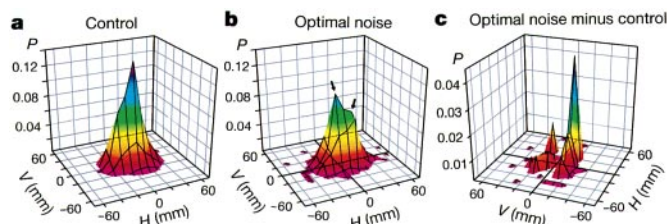


Figure 4 Probability distributions for strike locations. $P_s(x, y)$ values in controls (**a**) or during optimal noise (**b**) were obtained by binning the scatterplots in Fig. 2a and b, respectively, and normalizing to the total numbers of strikes for each. H , V , horizontal and vertical distance. The intersection of the axes corresponds to the midline centre axis of the fish's rostrum, as in Fig. 2. **c**, Subtracting these distributions (**b** minus **a**) revealed the region of space around the rostrum where optimal noise caused an increase in the strike probability. Only positive differences are shown: $P_s(x, y)_{\text{noise}} > P_s(x, y)_{\text{control}}$.

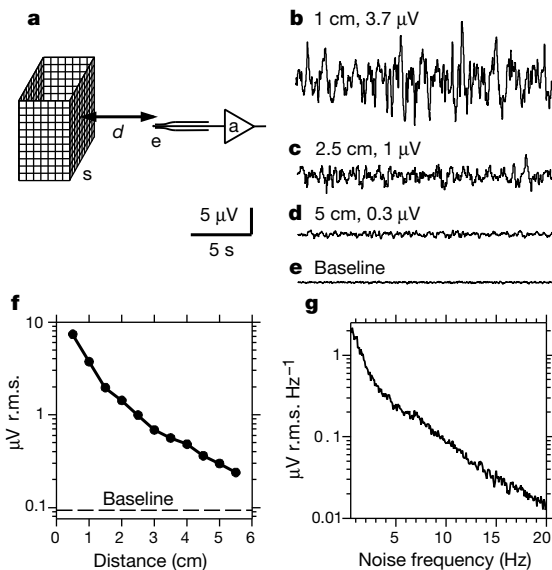


Figure 5 Electrical noise from *Daphnia* populations. **a**, An enclosure with sides of plastic screen (s), $5 \times 5 \times 3.3$ cm, contained 200 *Daphnia*, of ~ 3 mm body length. The side nearest a movable recording electrode (e) was coated with agarose to block water currents. **d**, horizontal distance. **a**, amplifier. The reference electrode was distant, in $760 \mu\text{S cm}^{-1}$ water. **b–d**, Sample recordings of *Daphnia* noise, at the distances listed, with r.m.s. amplitudes for the 0.5–20 Hz bandwidth that paddlefish electroreceptors are most sensitive to¹³. **e**, Control recording when plankton were absent, showing the electrode + instrumentation baseline (0.094 μV r.m.s.). **f**, Exponential falloff of r.m.s. amplitude in the 0.5–20 Hz band at distances beyond 3 cm. **g**, Amplitude spectrum of the swarm electrical noise at a distance of 10 mm, averaged over 54 windows, after baseline subtraction.

plankton at greater distances, increasing prey availability. The possibility that neural pathways from the brain might modulate the sensitivity of peripheral electroreceptors is unlikely, as such pathways have not been found²⁵. The optimal noise amplitude, 0.5–0.7 $\mu\text{V cm}^{-1}$, equivalent to a current density of 0.38–0.53 nA cm^{-2} , is close to the reported behavioural limit of sensitivity of ampullary electroreceptors in other freshwater fish¹⁶ ($\sim 1 \mu\text{V cm}^{-1}$). That these two numbers are comparable is expected for stochastic resonance: noise of comparable amplitude to a weak signal will enhance threshold crossings at a detector^{1–4}. Behavioural stochastic resonance may be an advantageous approach for measuring the sensitivity limits of sensory systems because, in contrast to the 50% behavioural response criterion used conventionally, stochastic resonance generates a well defined peak near the ‘threshold’.

Methods

Fish, swim mill and plankton

Data were from four juvenile paddlefish, 3–5 months old, 18–20 cm total length, deprived of food for 1–2 days before use. Experiments were run at night, when young paddlefish feed more briskly. The fish fed on individual *Daphnia* one-by-one in the viewing chamber of a ‘swim mill’^{10,13}, an apparatus for recirculating a stream of water, in a closed environmental room with only invisible near-infrared illumination. Paddlefish swim forward continuously, are ram ventilators¹⁰ and instinctively orient to swim into the oncoming stream. The water velocity ($\sim 10 \text{ cm s}^{-1}$) was adjusted to match normal swimming velocity, so a fish remained stationary relative to two co-aligned infrared video cameras which viewed the fish from the side and below (using a mirror), whose images were combined and videotaped¹³. An upstream collimator gave laminar flow, sweeping *Daphnia* in straight lines across the viewing chamber, parallel to the long axis of the fish’s rostrum. Controlled parameters included the water conductivity ($760 \mu\text{S cm}^{-1}$ at $21\text{--}22^\circ\text{C}$), the low density of plankton in the water (kept at ~ 2 per litre by adding aliquots of 10 as a fish consumed them, through tubing from an adjacent room, and monitoring the density during experiments using a special particle counter), and the *Daphnia* size (2.64 ± 0.18 mm). Control studies showed that the distribution of plankton across a transverse section of the viewing chamber was uniform except near the walls.

Applied noise

Noise from a General Radio 1390B generator was low-pass filtered by an 8-pole Bessel filter set to 50 Hz, attenuated, and then commanded a constant-current linear isolation unit driving the stimulus electrodes (Fig. 1a). A third plate electrode grounded the water. The noise amplitude was monitored during experiments by recording differentially between the stimulus electrodes. The optimal noise amplitude was verified offline using two separate movable focal horizontal recording electrodes, one 30 cm downstream from the other, while applying noise between the stimulus plates. The same approach verified less than 3% variation in r.m.s. stimulus amplitude across a transverse section of the fish viewing chamber.

Protocol

After one fish was transferred to the swim mill for 10 min, *Daphnia* were added and the fish allowed to feed for 15 min, giving a total habituation time of 25 min before data were collected. Continuous video recording was then begun, divided into 8-min epochs throughout the 48–80 min of data collection, always alternating between an 8-min control epoch (no noise), and an 8-min presentation of a noisy electrical current, whose amplitude could be selected. Results for an 8-min noise presentation were always normalized to results in the next 8-min control epoch, to compensate for any nonstationarity. The order of noise amplitudes was randomized using random numbers. To present one entire sequence of noise amplitudes to a fish required 2–4 experimental sessions, as each session was restricted to less than 95 min of feeding to limit satiation. Each of the four fish received two complete sequences of noise amplitudes. One of the fish received additional noise presentations at amplitudes around the optimum.

Measurement of videotaped strike locations

The analysis for Figs 2–4 was ‘blinded’ to avoid bias: assistants who were unaware of the experimental protocol analysed the tapes from start to end. Feeding strikes were identified by the fish opening its mouth and lunging at an approaching *Daphnia*, usually capturing it. The strike time on the videotape was noted. Then the videotape was backed up in pause mode until the *Daphnia* was in a vertical plane at the rostrum tip¹³ (see *x* and *y* arrows in Fig. 1a), a point chosen to be before the fish had reacted. If the fish’s rostrum was turned less than 15° from straight ahead¹³, ensuring valid fish-to-plankton distance measurements because *Daphnia* approached along paths nearly parallel to the rostrum’s long axis, the image was digitized. A software cursor was used to measure the location of the *Daphnia* and the rostrum tip centre in both camera views. Subtracting these gave the relative pixel separation from the fish to the *Daphnia*, both horizontally and vertically. Pixel distances were converted to millimetres using the image of a 10-cm spatial calibration bar, and corrected for camera perspective error. We measured 4,891 strikes for Fig. 3c and other parts of Figs 2–4, an unbiased 73% sample (the other strikes were not measured because fish were turned more than 15°). For Mood tests, the median vertical distance was made equal for noise and control groups¹⁸.

Received 20 July; accepted 14 September 1999.

- Wiesenfeld, K. & Moss, F. Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS. *Nature* **373**, 33–36 (1995).
- Gammaitoni, L., Hanggi, P., Jung, P. & Marchesoni, F. Stochastic resonance. *Rev. Mod. Phys.* **70**, 223–288 (1998).
- Bulsara, A. & Gammaitoni, L. Tuning in to noise. *Phys. Today* **49**, 39–45 (1996).
- Bezrukov, S. M. & Vodyanov, I. Stochastic resonance in non-dynamical systems without response thresholds. *Nature* **385**, 319–321 (1997).
- Douglass, J. K., Wilkens, L., Pantazazelou, E. & Moss, F. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature* **365**, 337–340 (1993).
- Levin, J. E. & Miller, J. P. Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature* **380**, 165–168 (1996).
- Collins, J. J., Imhoff, T. T. & Grigg, P. Noise-enhanced information transmission in rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance. *J. Neurophysiol.* **76**, 642–645 (1996).
- Collins, J. J., Imhoff, T. T. & Grigg, P. Noise-enhanced tactile sensation. *Nature* **383**, 770 (1996).
- Ruelle, R. & Hudson, P. L. Paddlefish (*Polyodon spathula*): growth and food of young of the year and a suggested technique for measuring length. *Trans. Am. Fisheries Soc.* **106**, 609–613 (1977).
- Burggren, W. W. & Bemis, W. E. Metabolism and ram gill ventilation in juvenile paddlefish, *Polyodon spathula* (Chondrostei: Polyodontidae). *Physiol. Zool.* **65**, 515–539 (1992).
- Jørgensen, J. M., Flock, Å. & Wersäll, J. The Lorenzian ampullae of *Polyodon spathula*. *Z. Zellforsch. Mikrosk. Anat.* **130**, 362–377 (1972).
- New, J. G. & Bodznick, D. Segregation of electroreceptive and mechanoreceptive lateral line afferents in the hindbrain of chondrosteian fishes. *Brain Res.* **336**, 89–98 (1985).
- Wilkens, L. A., Russell, D. F., Pei, X. & Gurgens, C. The paddlefish rostrum functions as an electrosensory antenna in plankton feeding. *Proc. R. Soc. Lond. B* **264**, 1723–1729 (1997).
- Jaramillo, F. & Wiesenfeld, K. Mechano-electrical transduction assisted by Brownian motion: a role for noise in the auditory system. *Nature Neurosci.* **1**, 384–388 (1998).
- Teeter, J. H., Szamier, R. B. & Bennett, M. V. L. Ampullary electroreceptors in the sturgeon *Scaphirhynchus platyrhynchus* (Rafinesque). *J. Comp. Physiol.* **138**, 213–223 (1980).
- Bullock, T. H. Electroreception. *Annu. Rev. Neurosci.* **5**, 121–170 (1982).
- Tricas, T. C. & New, J. G. Sensitivity and response dynamics of elasmobranch electrosensory primary afferent neurons to near threshold fields. *J. Comp. Physiol. A* **182**, 89–101 (1998).
- Marascuilo, L. A. & McSweeney, M. *Nonparametric and Distribution-Free Methods for the Social Sciences* 290–291 (Brooks/Cole, Monterey, 1977).
- Gibbons, J. D. *Nonparametric Methods for Quantitative Analysis* 223–225 (Holt, Rinehart & Winston, New York, 1976).
- Boucherle, M. M. & Frederick, V. R. *Daphnia* swarm in the harbor at Put-In-Bay. *Ohio J. Sci.* **56**, 90–91 (1976).
- Robbins, T. W. & Everitt, B. J. In *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 703–720 (MIT Press, Cambridge, Massachusetts, 1995).

22. Laming, P. R. in *Brain Mechanisms of Behaviour in Lower Vertebrates* (ed. Laming, P. R.) 203–222 (Cambridge Univ. Press, Cambridge, 1981).
23. Stemmler, M., Usher, M. & Niebur, E. Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. *Science* **269**, 1877–1880 (1995).
24. Gluckman, B. J. *et al.* Stochastic resonance in a neuronal network from mammalian brain. *Phys. Rev. Lett.* **77**, 4098–4101 (1996).
25. Bullock, T. H. in *Electroreception* (eds Bullock, T. H. & Heiligenberg, W.) 651–674 (Wiley, New York, 1986).
26. Pough, F. H., Heiser, J. B. & McFarland, W. N. *Vertebrate Life* 4th edn, 258 (Prentice Hall, Upper Saddle River, 1996).

Acknowledgements

This work was supported by the US Office of Naval Research, Physics Division (F.M.) and by the Whitehall Foundation (D.F.R.). Paddlefish were provided by the Missouri Department of Conservation. E. Wagner participated in preliminary experiments. We thank B. Wetrting, A. Tucker, J. Bingham, N. Tafra and D. Prescott for measuring the video data and culturing *Daphnia*, A. Fields for fish maintenance, W. Garver for constructing the plankton counter, and G. V. Welland, A. Neiman, J. Twitty and K. Dolan for suggestions. D.F.R. conceived of and designed the noise experiments, and carried them out and wrote the paper with F.M., who designed the analysis of Fig. 4. L.A.W. originated the paddlefish project here, the swim mill, and the scatterplot analysis.

Correspondence and requests for materials should be addressed to D.F.R. (e-mail: drussell@admiral.umsl.edu).

The amygdala modulates prefrontal cortex activity relative to conditioned fear

René Garcia⁺, Rose-Marie Vouimba[†], Michel Baudry[†] & Richard F. Thompson[†]

⁺ Laboratoire de Neurosciences Cognitives, CNRS UMR 5807, Université de Bordeaux I, Avenue des Facultés, 33405 Talence, France
[†] Neuroscience Program, University of Southern California, Los Angeles, California 90089-2520, USA

Animals learn that a tone can predict the occurrence of an electric shock through classical conditioning. Mice or rats trained in this manner display fear responses, such as freezing behaviour, when they hear the conditioned tone. Studies using amygdalotomized rats have shown that the amygdala is required for both the acquisition and expression of learned fear responses^{1–3}. Freezing to a conditioned tone is enhanced following damage to the dorsal part of the medial prefrontal cortex⁴, indicating that this area may be involved in fear reduction. Here we show that prefrontal neurons reduce their spontaneous activity in the presence of a conditioned aversive tone as a function of the degree of fear. The depression in prefrontal spontaneous activity is related to amygdala activity but not to the freezing response itself. These data indicate that, in the presence of threatening stimuli, the amygdala controls both fear expression and prefrontal neuronal activity. They suggest that abnormal amygdala-induced modulation of prefrontal neuronal activity may be involved in the pathophysiology of certain forms of anxiety disorder.

To test whether neurons in the medial prefrontal cortex are involved in mechanisms of fear modulation, mice were implanted with two recording electrodes in the dorsal part of this cortical area. The whole experiment comprised four phases, each separated by a 3-day period: habituation phase (5 days), first training phase (2 days), second training phase (5 days) and test phase (1 day). The habituation and test phases took place in a recording chamber with behavioural and electrophysiological equipment for assessing freezing and recording multi-unit activity, respectively. The training phase took place in a different chamber. We used two conditioned stimuli: conditioned fear excitation (CS: 20-s tone) and conditioned

fear inhibition (CI: 20-s light). An electric shock was the unconditioned stimulus (US). In the first phase of training, five CS–US pairings were presented with a variable intertrial interval (60–180 s) on each of two consecutive days. Following this fear-conditioning phase, animals were divided into two groups (CI–CS paired and CI–CS unpaired groups). All animals underwent a second training phase in which five CS–US pairings were randomly mixed with five additional trials, where either the CS was preceded immediately by the CI and the US was withheld (CI–CS paired group) or the CI was presented alone (CI–CS unpaired group). The intertrial interval ranged from 60–180 s. Animals in the CI–CS paired group were then trained according to a Pavlovian conditioned inhibition procedure⁵ that reduces behavioural changes normally attributed to fear excitation. For example, it has been reported that, following conditioning (light–shock and tone–light compounds), rats displayed less fear-potentiated startle to the light CS when it was preceded by the tone CI⁶. We hypothesized that animals in the CI–CS paired group would display less freezing than animals in the CI–CS unpaired group, which is considered to be a good control group for conditioned inhibition⁷.

To test the effects of both CI and CS on spontaneous neuronal activity in the medial prefrontal cortex, multi-unit recordings were performed in the recording chamber on the last day of the experiment

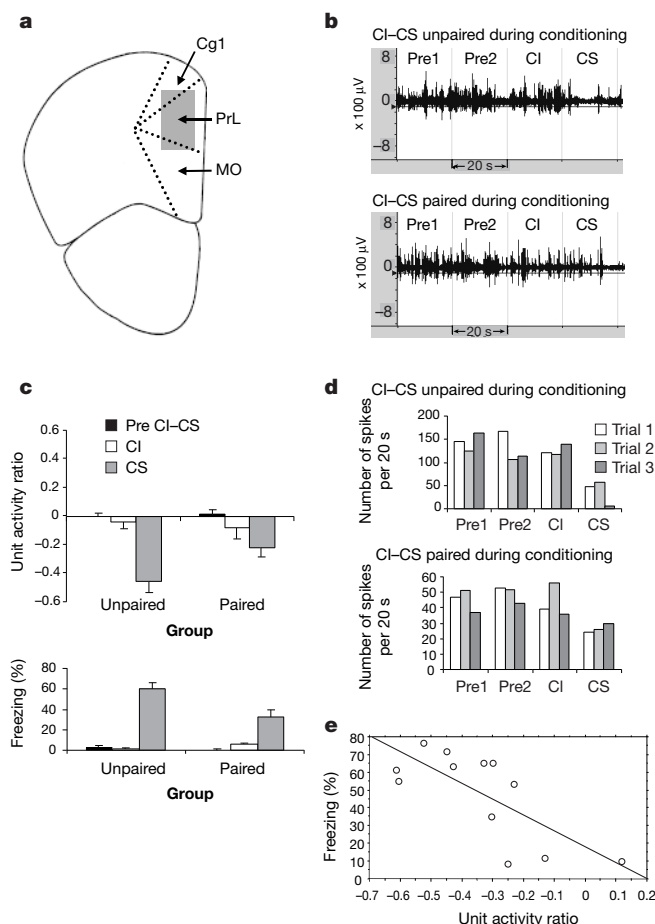


Figure 1 Conditioned inhibition: electrophysiology and behaviour. **a**, Electrode placements in the prefrontal cortex (grey area) Cg1: cingulate area 1; PrL and MO: prefrontal and medial orbital cortices, respectively³⁰. **b**, **d**, Representative changes in activity before (Pre 1 and 2) and during CI–CS presentation. **c**, Top, mean unit activity ratios. CS presentation caused a greater decrease in prefrontal activity in unpaired animals than in paired animals ($F_{1,10} = 5.9$; $P < 0.05$). Bottom, mean percentage freezing. Unpaired mice displayed more freezing than paired mice ($F_{1,10} = 7.77$; $P < 0.02$). **e**, Relationship between behavioural and electrophysiological data from all 12 animals ($r = 0.73$; $P = 0.0063$).

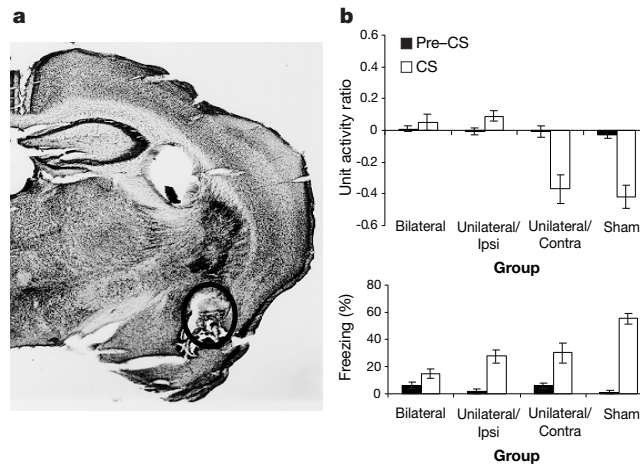


Figure 2 Effect of amygdala lesion. **a**, Representative photomicrograph of a coronal section showing lesion (circle) through the basolateral amygdala. **b**, Top, mean unit activity ratios. During CS presentation, unilateral/contra animals displayed a significant decrease in prefrontal activity compared to both bilateral and unilateral/ipsi animals (both

$P < 0.05$), but did not differ from sham animals. Bottom, mean percentage freezing. In the bilateral group, CS presentation did not induce any significant freezing. In the presence of the CS, animals with unilateral lesions displayed significantly more freezing than the bilateral animals ($P < 0.05$), but less freezing than the sham animals ($P < 0.05$).

(test phase). Prefrontal electrical activity was accepted for recordings only if the signal-to-noise ratio was at least 4:1 for the largest spikes. Under these conditions, the activities of at least two units were discriminable by amplitude window. Three CI-CS pairings were presented with an intertrial interval of 60–180 s. Multi-unit analysis was performed as described⁸, and changes in firing were quantified according to the formula $(B - A)/(B + A)$, where B is the total number of spikes during the CS or CI presentation and A is the total number of spikes during the pre-CI-CS period. The same formula was used to determine the stability of the response during the pre-CI-CS period.

Examination of counterstained brain sections revealed that electrode tips were all located in the prelimbic cortex (Fig. 1a). The spontaneous firing rates of neurons in these areas ranged from 1 to 8 Hz and were stable during the pre-CI-CS period (Fig. 1d). In response to CS, neuronal activity was exclusively inhibited (Fig. 1b, d). A similar decrease in excitability in the presence of a conditioned stimulator of fear occurs in the lateral septum⁸. However, in contrast to septal neurons, which display an opposite pattern of spontaneous activity in the presence of a conditioned inhibitor of fear (an increase in activity⁸), prefrontal neurons did not present any significant change in response to CI presentation. Nevertheless, the decrease in activity observed with conditioned fear excitation was significantly reduced in animals submitted to conditioned inhibition (CI-CS paired group, Fig. 1b, c).

The amount of freezing time was measured during the 20 s immediately before the CI onset (pre-CI-CS period) and during both the 20-s CI and 20-s CS presentations. After conditioning, mice in the CI-CS unpaired group displayed a marked freezing response as compared to freezing level during the pre CI-CS period. Mice in the CI-CS paired group displayed less freezing than mice in the CI-CS unpaired group, indicating the acquisition of conditioned inhibition (Fig. 1c).

These electrophysiological and behavioural data indicate that the level of CS-induced depression in prefrontal spontaneous activity is related to the level of CS-evoked freezing behaviour (Fig. 1e). If this were true, a strong reduction in conditioned freezing produced by postconditioning amygdectomy should abolish CS-induced changes in prefrontal neuronal activity. In addition, as the amygdala from one side of the brain projects poorly to the contralateral prefrontal cortex⁹ and as unilateral amygdectomy reduces but does not abolish conditioned freezing¹⁰, such a lesion should block CS-induced depression in the ipsilateral prefrontal cortex while leaving intact CS-induced depression in the contralateral prefrontal

cortex. To test this, mice were implanted with recording electrodes in medial prefrontal cortex and lesioning electrodes in the basolateral complex. Animals were submitted only to the first training phase (tone-shock association) because it was assumed that, if amygdectomy were to suppress CS-induced depression in prefrontal spontaneous activity in the presence of the CS alone, it would also do so following conditioned inhibition of fear (when the CS is preceded by the CI). Thus, animals received, 30 min after the tone-shock association, either a bilateral lesion (bilateral lesion group), a unilateral lesion (unilateral/ipsi group, with recording electrodes in the ipsilateral prefrontal cortex; unilateral/contra group, with recording electrodes in the contralateral prefrontal cortex), or no lesion of the basolateral complex (sham lesion group). The auditory test was performed 3 days later.

Figure 2a shows a representative photomicrograph of the amygdala lesion. Only animals with large electrolytic lesions including the basolateral nucleus of the amygdala were considered for data analyses ($n = 18$). Electrophysiological data show that, in the presence of the conditioned tone, animals in the sham lesioned group ($n = 6$) displayed a strong decrease in prefrontal multi-unit activity (Fig. 2b). This effect was completely suppressed in animals from the bilateral lesion group ($n = 6$). However, in the unilateral lesion groups ($n = 12$), prefrontal recordings from the contralateral side to the lesion (unilateral/contra; $n = 6$) were characterized by a CS-induced decrease in prefrontal activity similar to that observed in sham lesioned animals. Recordings performed on the side ipsilateral to the lesion (unilateral/ipsi; $n = 6$) indicated a lack of CS-induced depression in prefrontal activity. No animals exhibited freezing behaviour in the pre-CS period. However, during the CS presentation, animals with bilateral amygdala lesions froze less than animals with unilateral lesions, which in turn displayed less freezing behaviour than animals from the sham lesioned group (Fig. 2b). Thus, postconditioning amygdectomy led to behavioural results similar to those reported¹⁰ for preconditioning amygdectomy.

Our findings provide direct evidence for the control by the basolateral amygdala of learned auditory fear-induced changes in neuronal activity in the medial prefrontal cortex. These data fit well with and account for numerous pieces of information. First, anatomical and electrophysiological studies provide evidence for auditory connections in the basolateral complex of the amygdala^{11–13}. Second, both unconditioned and conditioned stimuli induce long-term potentiation of auditory-evoked synaptic transmission in the basolateral complex of the amygdala^{14–16}. Third, stimulation of the basolateral nucleus predominantly induces

inhibitory responses in the medial prefrontal cortex¹⁷. Collectively, these observations indicate that detection of a conditioned aversive auditory stimulus may activate neurons in the basolateral complex of the amygdala, which may in turn control both the expression of learned fear and electrical activity in the medial prefrontal cortex (characterized by an inhibitory effect). As the direct projection from the basolateral nucleus of the amygdala to the medial prefrontal cortex involves pyramidal neurons which contain excitatory neurotransmitters¹⁸, the indirect brain circuit by which the amygdala exerts its inhibitory effect on prefrontal neuronal activity remains to be determined. One of the putative inhibitory pathways may be the amygdala–brainstem–prefrontal cortex connection. Indeed, the amygdala sends fibres to the brainstem dopaminergic ventral tegmental area¹⁹ which, in turn, projects to prefrontal cortex^{20,21}. Moreover, (i) threatening stimuli activate the dopaminergic system in the medial prefrontal cortex²², (ii) amygdala lesions suppress this activation^{22,23} and (iii) activation of the dopaminergic system decreases prefrontal spontaneous activity^{24,25}. Thus, the amygdala may exert its influence (inhibitory effect on prefrontal spontaneous activity) through activation of the mesocortical dopaminergic system. Alternatively, the amygdala may also exert its influence through other extracortical pathways and systems or directly through intracortical networks²⁶.

The functional role of amygdala activity-induced inhibition of prefrontal cortex activity also remains to be determined. However, it is known both in humans²⁷ and in animals⁴, that, unlike the amygdala, the medial prefrontal cortex is not necessary to represent the affective attributes of an explicit CS. Taking into account the role of the prefrontal cortex in cognitive functions and the fact that lesion of the prefrontal cortex retards fear extinction⁴, we suggest that CS-induced prefrontal neuronal plasticity may be related to representation of how well the CS predicts danger. More precisely, when danger is well predicted by the CS, prefrontal neuronal activity is strongly depressed (which may encode the strong certainty of the occurrence of danger). However, when the predictive ability of the CS for danger is weakened, for instance by a conditioned inhibition procedure (as here) or an extinction procedure (repeated presentations of the CS alone)²⁸, there is weak depression (as here) and even potentiation²⁸ of prefrontal neuronal activity (which may encode the low certainty of the occurrence of danger). Therefore, as shown for decision-making²⁷ or for expected outcomes²⁹, the integrity of the amygdala appears to be necessary for the formation of a representation of the degree of predictability of an aversive event in the prefrontal cortex. At a more clinical level, we suggest that abnormal amygdala-induced modulation of prefrontal neuronal activity may be involved in the pathophysiology of certain forms of anxiety disorder. □

Methods

Electrode implantation, electrophysiology and lesion

We used 36 male mice (C57BL/6 JI Co; 4–6 months; 28–32 g) with electrodes implanted under anaesthetic (a mixture of ketamine at 80 mg kg⁻¹ and xylazine at 20 mg kg⁻¹). Electrodes were made of platinum–iridium wires (50 µm in diameter) and were positioned in the dorsal part of the medial prefrontal cortex for multi-unit activity recording and in the basolateral complex for postconditioning lesion. Three days after surgery, animals were handled for 1 min and placed in a Plexiglas recording chamber (17 × 17 × 20 cm high with a plastic floor) for 10 min on five consecutive days. The recording chamber was washed with 1% acetic acid before and after each session. The behaviour of the mice was observed using a video camera monitor system. Spontaneous multi-unit activity was recorded through JFET operational amplifiers connected to the mouse's head to minimize artefacts due to head movement. Microcables from the JFET were relayed at the top of the recording chamber by a multichannel rotating connector. This system allowed the animal free movement within the recording chamber. Prefrontal neuronal activity was sent to a differential band-pass amplifier (0.3–10 kHz; gain × 10,000) and a discriminator set to count only the larger spikes (Axoscope, Axon Instruments). For lesioning, an anodal current (1 mA) was passed through the basolateral electrode for 20 s with the mice anaesthetized. Unlesioned mice were also anaesthetized but no current was passed through the amygdala electrode. The exact placement of both the recording and lesioning electrodes was verified at the end of the experiments by standard histological methods.

Conditioning

Computerized training (Forth software, USC) took place in a semi-opaque box (20 × 20 × 25 cm high) with a shock grid made of stainless steel rods (0.45 cm in diameter, spaced 1.9 cm apart) and a removable 12 W light bulb at the top of the box. A speaker providing a 1 kHz, 80 dB tone was located 15 cm in front of the box. The shock grid was connected to a shock generator and scrambler to provide a 0.5 s, 0.5 mA footshock. The conditioning box and the floor were cleaned with 70% ethanol before and after each session. A background noise of 55 dB was provided throughout all experiments.

Received 21 July; accepted 8 September 1999.

- Davis, M. in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J. P.) 255–305 (Wiley-Liss, New York, 1992).
- LeDoux, J. E. Emotion: clues from brain. *Annu. Rev. Psychol.* **46**, 209–235 (1995).
- Maren, S. & Fanselow, M. S. The amygdala and fear conditioning: has the nut been cracked? *Neuron* **16**, 237–240 (1996).
- Morgan, M. A. & LeDoux, J. E. Differential contribution of dorsal and ventral medial prefrontal cortex to acquisition and extinction of conditioned fear in rats. *Behav. Neurosci.* **109**, 681–688 (1995).
- Rescorla, R. A. Pavlovian conditioned inhibition. *Psychol. Bull.* **72**, 77–94 (1969).
- Campeau, S. *et al.* Elicitation and reduction of fear: behavioural and neuroendocrine indices and brain induction of the immediate-early gene c-fos. *Neuroscience* **78**, 1087–1104 (1997).
- Papini, M. R. & Bitterman, M. E. The two-test strategy in the study of inhibitory conditioning. *J. Exp. Psychol. Anim. Behav. Process.* **19**, 342–352 (1993).
- Thomas, E. & Yadin, E. Multiple unit activity in the septum during Pavlovian aversive conditioning: evidence for an inhibitory role for the septum. *Exp. Neurol.* **69**, 50–60 (1980).
- McDonald, A. J. Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience* **44**, 1–14 (1991).
- LaBar, K. S. & LeDoux, J. E. Partial disruption of fear conditioning in rats with unilateral amygdala damage: correspondence with unilateral temporal lobectomy in humans. *Behav. Neurosci.* **110**, 991–997 (1996).
- Farb, C. R. & LeDoux, J. E. NMDA and AMPA receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse* **27**, 106–121 (1997).
- LeDoux, J. E., Farb, C. & Ruggiero, D. A. Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *J. Neurosci.* **10**, 1043–1054 (1990).
- Turner, B. H. & Herkenham, M. Thalamo-amygdaloid projections in the rat: a test of the amygdala's role in sensory processing. *J. Comp. Neurol.* **313**, 295–325 (1991).
- Garcia, R., Paquereau, J., Vouimba, R. M. & Jaffard, R. Conditioned fear stress enhances auditory-evoked potentials in the basolateral amygdala of the awake rat. *Eur. J. Neurosci. (Suppl.)* **9**, 194 (1996).
- Garcia, R., Paquereau, J., Vouimba, R. M. & Jaffard, R. Footshock stress but not contextual fear conditioning induces long-term enhancement of auditory-evoked potentials in the basolateral amygdala of the freely behaving rat. *Eur. J. Neurosci.* **10**, 457–463 (1998).
- Rogan, M. T., Staubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
- Pérez-Jaranay, J. M. & Vives, F. Electrophysiological study of the response of medial prefrontal cortex neurons to stimulation of the basolateral nucleus of the amygdala in the rat. *Brain Res.* **564**, 97–101 (1991).
- McDonald, A. J. Organization of amygdaloid projections to the mediodorsal thalamus and prefrontal cortex: a fluorescence retrograde transport study in the rat. *J. Comp. Neurol.* **262**, 46–58 (1987).
- Gonzales, C. & Chesselet, M. F. Amygdalo-nigral pathways: an anterograde study in the rat with phaseolus vulgaris leucoagglutinin (PHAL-L). *J. Comp. Neurol.* **297**, 182–200 (1990).
- Lindvall, O., Björklund, A., Moore, R. Y. & Stenevi, U. Mesencephalic dopamine neurons projecting to neocortex. *Brain Res.* **81**, 325–331 (1974).
- Fuxe, K., Hökfelt, T., Johansson, O., Lidbrink, P. & Ljungdahl, A. The origin of the dopamine nerve terminals in limbic and frontal cortex. Evidence for meso-cortico dopamine neurons. *Brain Res.* **82**, 349–355 (1974).
- Davis, M. *et al.* Stress-induced activation of prefrontal cortex dopamine turnover: blockade by lesions of the amygdala. *Brain Res.* **664**, 207–210 (1994).
- Goldstein, L. E., Rasmussen, A. M., Bunney, B. S. & Roth, R. H. Role of the amygdala in the coordination of behavioral, neuroendocrine, and prefrontal cortical monoamine responses to psychological stress in the rat. *J. Neurosci.* **16**, 4787–4798 (1996).
- Ferron, A., Thierry, A. M., Le Douarin, C. & Glowinski, J. Inhibitory influence of the mesocortical dopaminergic system on spontaneous activity or excitatory response induced from the thalamic mediodorsal nucleus in the rat medial prefrontal cortex. *Brain Res.* **302**, 257–265 (1984).
- Mantz, J., Milla, C., Glowinski, J. & Thierry, A. M. Differential effects of ascending neurons containing dopamine and noradrenaline in the control of spontaneous activity and of evoked responses in the rat prefrontal cortex. *Neuroscience* **27**, 517–526 (1988).
- Schoenbaum, G., Chiba, A. A. & Gallagher, M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* **19**, 1876–1884 (1999).
- Bechara, A., Damasio, H., Damasio, A. R. & Lee, G. P. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* **19**, 5473–5481 (1999).
- Herry, C., Vouimba, R. M. & Garcia, R. Plasticity in the thalamo-prefrontal cortical transmission in behaving mice. *J. Neurophysiol.* (in the press).
- Schoenbaum, G., Chiba, A. A. & Gallagher, M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neurosci.* **1**, 155–159 (1998).
- Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* (Academic, San Diego, 1997).

Acknowledgements

This work was supported by grants from the Fondation Fyssen (R.M.V.) and from NATO (R.G.).

Correspondence and requests for materials should be addressed to R.G. (e-mail: garcia@neurocog.u-bordeaux.fr).

Kainate receptors are involved in synaptic plasticity

Zuner A. Bortolotto*, Vernon R. J. Clarke*, Caroline M. Delany*, Michael C. Parry*, Ilse Smolders*†, Michel Vignes*‡, Ken H. Ho§, Peter Miu§, Bradford T. Brinton§, Robert Fantáske§, Ann Ogdén||, Mary Gates||, Paul L. Ornstein||, David Lodge||, David Bleakman|| & Graham L. Collingridge*

* MRC Centre for Synaptic Plasticity, Department of Anatomy, Medical School, University of Bristol, Bristol, BS8 1TD, UK

† Department of Pharmaceutical Chemistry & Drug Analysis, Pharmaceutical Institute, Vrije Universiteit Brussels (VUB), Laarbeeklaan 103, 1090 Brussels, Belgium

‡ Laboratoire 'Plasticité Cérébrale', EP 628 CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

§ Allelix Biopharmaceuticals, 6850 Goreway Drive, Mississauga, Ontario L4V 1V7, Canada

|| Eli Lilly & Co., Lilly Corporate Centre, Indianapolis 46285, Indiana, USA

The ability of synapses to modify their synaptic strength in response to activity is a fundamental property of the nervous system and may be an essential component of learning and memory¹. There are three classes of ionotropic glutamate receptor, namely NMDA (N-methyl-D-aspartate), AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole-4-propionic acid) and kainate

receptors²; critical roles in synaptic plasticity have been identified for two of these. Thus, at many synapses in the brain, transient activation of NMDA receptors leads to a persistent modification in the strength of synaptic transmission mediated by AMPA receptors^{3,4}. Here, to determine whether kainate receptors⁵⁻⁷ are involved in synaptic plasticity, we have used a new antagonist, LY382884 ((3S, 4aR, 6S, 8aR)-6-((4-carboxyphenyl)methyl)-1,2,3,4,4a,5,6,7,8,8a-decahydroisoquinoline-3-carboxylic acid), which antagonizes kainate receptors at concentrations that do not affect AMPA or NMDA receptors. We find that LY382884 is a selective antagonist at neuronal kainate receptors containing the GluR5 subunit. It has no effect on long-term potentiation (LTP) that is dependent on NMDA receptors but prevents the induction of mossy fibre LTP, which is independent of NMDA receptors. Thus, kainate receptors can act as the induction trigger for long-term changes in synaptic transmission.

We have previously described two compounds, LY293558 ((3S, 4aR, 6R, 8aR)-6-[2-(1(2H-tetrazol-5-yl)ethyl)-decahydroisoquinoline-3-carboxylic acid]⁸ and LY294486 ((3SR, 4aRS, 6SR, 8aRS)-6-(((1H-tetrazol-5-yl) methyl) oxy) methyl)-1,2,3,4,4a,5,6,7,8,8a-decahydroisoquinoline-3-carboxylic acid)⁹, that inhibit responses mediated by kainate receptors containing the GluR5 but not the GluR6 subunit. However, these kainate receptor antagonists are also potent AMPA receptor antagonists, limiting their use except under circumstances where synaptic transmission mediated by AMPA receptors is blocked. We therefore examined other members of this family of decahydroisoquinoline compounds using ligand-binding assays; one compound, LY382884, was identified as

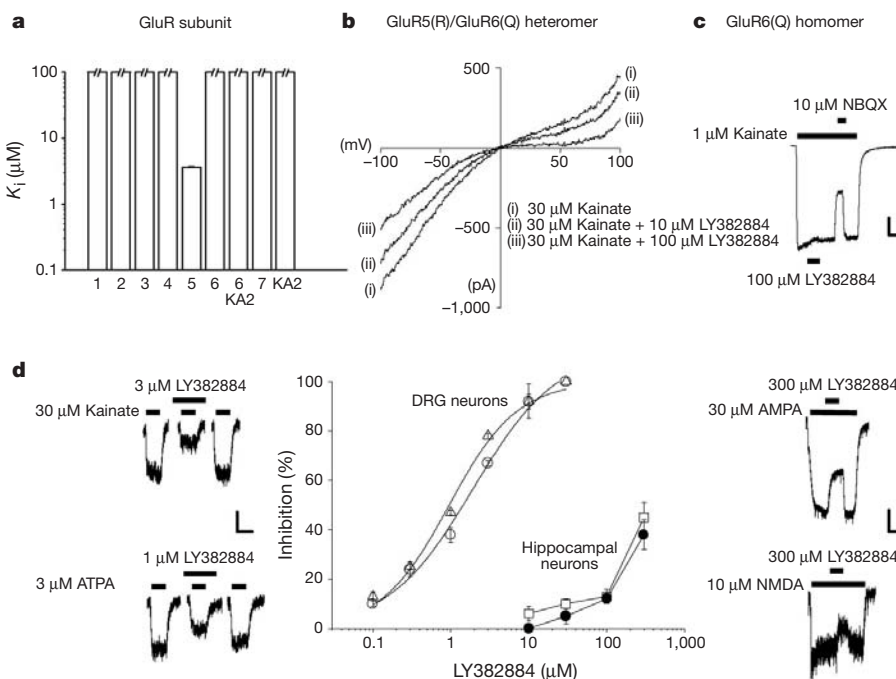


Figure 1 LY382884 is a selective GluR5 kainate receptor antagonist. **a**, Selective displacement of binding to human recombinant GluR5 receptors. K_i values were estimated from 11-point competition assays, from 3–4 separate preparations. Histograms with open tops indicate a binding affinity $>100 \mu\text{M}$ (displacements using $100 \mu\text{M}$ LY382884: GluR1, $23 \pm 5\%$; GluR2, $26 \pm 3\%$; GluR3, $5 \pm 3\%$; GluR4, $24 \pm 2\%$; GluR6, $1 \pm 1\%$; GluR7, $26 \pm 6\%$; KA2, $8 \pm 3\%$; GluR6 + KA2, $25 \pm 16\%$). **b**, LY382884 antagonizes GluR5/GluR6 heteromeric receptors. A voltage ramp was applied during steady-state responses to kainate ($30 \mu\text{M}$). Cells with a significant response generated by heteromeric receptors were characterized by less rectification than homomeric GluR6(Q) (the ratio of currents at $+50/-50 \text{ mV}$ was 0.66 ± 0.07 for heteromeric receptors ($n = 8$) and 0.06 ± 0.03 for homomeric GluR6(Q) ($n = 5$)). Kainate currents were inhibited by LY382884 (10 and $100 \mu\text{M}$) by 18 ± 3 and $41 \pm 3\%$, respectively (-50 mV), and by 30 ± 6 and $70 \pm$

6% , respectively ($+50 \text{ mV}$). The voltage-dependence of the antagonism was significant ($P < 0.01$, $n = 8$). **c**, LY382884 does not affect $1 \mu\text{M}$ kainate currents in cells expressing homomeric GluR6(Q) ($4 \pm 3\%$; $n = 3$). The effectiveness of NBQX ($10 \mu\text{M}$) is shown for comparison. Holding voltage (V_h) = -70 mV ; scale bar represents 100 pA vertical and 10 s horizontal. (LY382884 ($100 \mu\text{M}$) also had little effect on $30 \mu\text{M}$ kainate currents in GluR6(Q) homomers ($6 \pm 1\%$; $n = 5$)). **d**, Functional selectivity of LY382884 at native ionotropic glutamate receptors. The graph shows concentration-dependent antagonism by LY382884 of responses of DRG neurons to $3 \mu\text{M}$ ATPA (triangles) and $30 \mu\text{M}$ kainate (open circles) and of hippocampal neurons to $10 \mu\text{M}$ NMDA (filled circles) and $30 \mu\text{M}$ AMPA (squares) ($V_h = -70 \text{ mV}$). Data were obtained from at least three separate cells at each concentration. Scale bars for representative currents are 50 pA (kainate), 75 pA (ATPA), 200 pA (AMPA) and 50 pA (NMDA) (vertical) and 10 s (horizontal).

having considerably greater selectivity for GluR5 than for GluR2 (ref. 10). The ability of LY382884 to displace binding at a wide range of recombinant AMPA and kainate receptors is shown in Fig. 1a. Binding to GluR5 was displaced with an inhibition constant (K_i) of $4.0 \pm 0.2 \mu\text{M}$ ($n = 4$), whereas binding to GluR1–4, GluR6, GluR7, KA2 and a heteromeric assembly of GluR6 and KA2 was displaced with K_i values in excess of $100 \mu\text{M}$. The kainate receptor antagonist activity of LY382884 is shown for recombinant and native receptors in Fig. 1b–d. LY382884 antagonized kainate-induced currents in GluR5/GluR6 heteromers (Fig. 1b) but not in GluR6 homomers (Fig. 1c). The heteromer was made by co-expressing GluR6(Q), which alone generated strongly rectifying currents, with GluR5(R), which alone was non-functional; this co-expression yielded a more linear I – V relationship than GluR6 alone, in response to voltage ramps between -100 and $+100$ mV. In cells displaying this more linear rectification, LY382884 dose dependently reduced the current induced by $30 \mu\text{M}$ kainate and increased rectification, presumably because a greater proportion of GluR6 homomers contributed to the residual current (Fig. 1b). The absence of functional antagonism at recombinant GluR6 kainate receptors is shown in Fig. 1c, which compares LY382884 with the AMPA/kainate receptor antagonist NBQX (2,3-dihydroxy-6-nitro-7-sulphamoylbenzo(f)quinoxaline). LY382884 had no observable antagonistic effects at human GluR6 kainate receptors at concentrations of up to $100 \mu\text{M}$. In rat dorsal root ganglion (DRG) neurons, which express GluR5 kainate receptors^{8,11,12}, LY382884 inhibited currents evoked by kainate ($30 \mu\text{M}$) and a selective GluR5 kainate receptor ligand, ATPA ((RS)-2-amino-3-(3-hydroxy-5-*tert*-butylisoxazol-4-yl)propanoic acid) ($3 \mu\text{M}$) in a concentration-dependent manner (half maximal concentration (IC_{50}) $0.95 \pm 0.16 \mu\text{M}$ ($n = 6$) and $1.19 \pm 0.79 \mu\text{M}$ ($n = 6$), respectively). We also established the antagonistic activity of LY382884 at hippocampal AMPA and NMDA receptors (Fig. 1d). LY382884 had little or no effect on currents evoked by AMPA ($30 \mu\text{M}$) or NMDA ($10 \mu\text{M}$) at a concentration of $10 \mu\text{M}$.

To determine whether LY382884 can be used to antagonize neuronal kainate receptors selectively in an intact slice preparation, we performed experiments on the CA1 region of the hippocampus. We recorded excitatory postsynaptic potentials (EPSPs) mediated by AMPA receptors intracellularly and determined their sensitivity by sequentially applying increasing concentrations of the antagonist. LY382884 depressed the synaptic AMPA receptor-mediated response with an IC_{50} of $87 \mu\text{M}$ ($n = 3$; Fig. 2a). In these experiments, $10 \mu\text{M}$ was the maximum concentration that could be used before AMPA receptor-mediated synaptic transmission was affected. Neither monosynaptic γ -aminobutyric acid (GABA_A) and GABA_B receptor-mediated synaptic transmission¹³ nor passive membrane properties were affected by $10 \mu\text{M}$ LY382884 ($n = 5$; data not shown). Next, we determined the effectiveness of $10 \mu\text{M}$ LY382884 as a kainate receptor antagonist by testing its ability to inhibit the depression of AMPA receptor-mediated synaptic transmission induced by ATPA, using field potential recordings in slices obtained from juvenile rats¹⁴. ATPA depressed field EPSPs (fEPSPs), at a concentration of $1 \mu\text{M}$ ($n = 5$) or $3 \mu\text{M}$ ($n = 6$), respectively, by 57 ± 6 and $63 \pm 6\%$ under control conditions but by only 6 ± 2 and $18 \pm 9\%$ in the presence of LY382884 (data not shown).

Given the high density of kainate receptors in area CA3 (ref. 15), we extended our analysis to this region. LY382884 similarly antagonized the depression of AMPA receptor-mediated EPSPs induced by ATPA. Thus, $1 \mu\text{M}$ ATPA depressed associational/commissural and mossy-fibre-evoked fEPSPs, respectively, by 34 ± 2 and $42 \pm 11\%$ under control conditions but by only 6 ± 4 and $10 \pm 5\%$ in the presence of $10 \mu\text{M}$ LY382884 ($n = 4$; Fig. 2b). Synaptic transmission in area CA3 is also inhibited by activation of metabotropic glutamate (mGlu) receptors, raising the possibility that ATPA and LY382884 are acting through these receptors. However, the mGlu receptor antagonist (S)- α -methyl-4-carboxyphenylglycine (MCPG) had no effect on ATPA-induced depression ($n = 5$), and LY382884

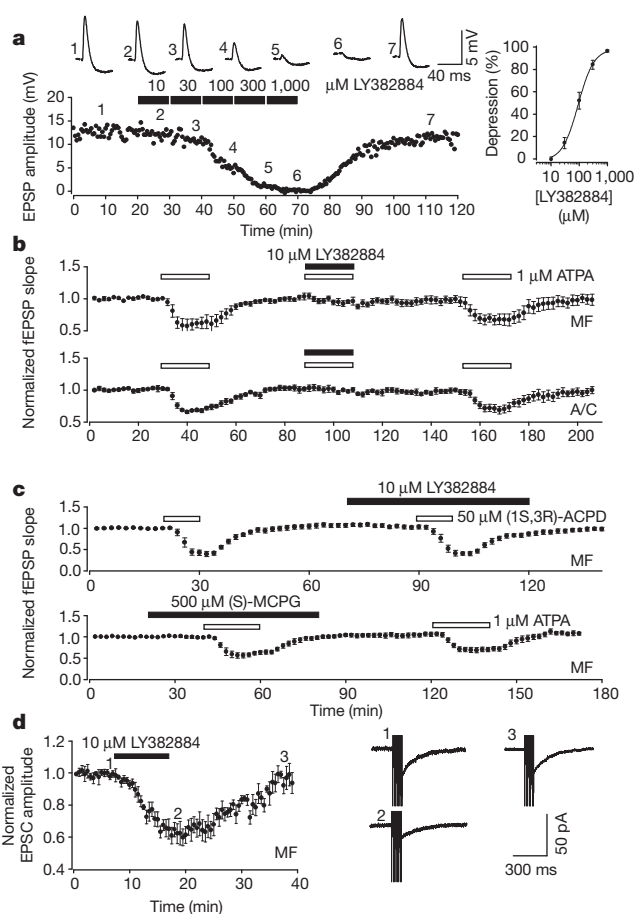


Figure 2 LY382884 as an antagonist in rat hippocampal slices. **a**, Concentration-dependent antagonism of AMPA receptor-mediated EPSPs. The left graph shows a single representative example and the right graph plots peak synaptic depression versus [LY382884] for three neurons ($\text{EC}_{50} = 87 \mu\text{M}$; neurons held at -73 ± 3 mV). **b**, Pooled data ($n = 4$) showing reversible antagonism by LY382884 of ATPA-induced depression of field EPSPs in area CA3, evoked alternatively by stimulation of mossy fibres (MF) and associational/commissural (A/C) inputs. **c**, Pooled data to show that LY382884 does not affect (1S,3R)-ACPD-induced depression ($n = 6$) and that MCPG does not affect ATPA-induced depression ($n = 5$) of mossy-fibre-evoked fEPSPs. **d**, Reversible antagonism by LY382884 of kainate receptor-mediated EPSCs, evoked by high-frequency stimulation (10 shocks at 100 Hz) of mossy fibres, in the presence of a cocktail of antagonists¹⁶ to prevent activation of AMPA, NMDA or GABA receptor-mediated synaptic currents ($n = 7$).

had no effect on depression induced by the mGlu receptor agonist (1S,3R)-1-aminocyclopentane-1,3-dicarboxylic acid ((1S,3R)-ACPD; $n = 6$; Fig. 2c). Furthermore, $10 \mu\text{M}$ LY382884 had no effect on depolarization of CA3 neurons induced by $30 \mu\text{M}$ (RS)-DHPG ((RS)-3,5-dihydroxyphenylglycine) (mean depolarizations before and after addition of LY382884, 10 ± 3 and 12 ± 2 mV, respectively; $n = 8$), precluding an effect of the antagonist on postsynaptic group I mGlu receptors. CA3 neurons possess postsynaptic kainate receptors that can be activated synaptically by brief, high frequency tetanus^{16,17}. LY382884 ($10 \mu\text{M}$) antagonized this kainate receptor-mediated excitatory postsynaptic current (EPSC), recorded under whole-cell voltage-clamp conditions at -70 mV, by $38 \pm 4\%$ ($n = 7$; Fig. 2d).

Having established that LY382884 is a selective kainate receptor antagonist, we wished to determine whether kainate receptors are involved in the induction of LTP at mossy fibre synapses, given that

LTP at these synapses is classically independent of NMDA receptors¹⁸. Mossy fibre LTP was completely prevented, in a reversible manner, by LY382884. (The mean potentiation 60 min after tetanization (100 Hz, 1 s, test intensity) in the presence and after washout of LY382884 was $1 \pm 4\%$ and $48 \pm 10\%$, respectively; $n = 7$; Fig. 3a, b.) The ability of LY382884 to block the induction of LTP was pathway specific, as NMDA receptor-dependent LTP in the CA3 region of the hippocampus, evoked by tetanic stimulation of associational/commissural fibres, was fully resistant to the actions of LY382884. The mean potentiation 60 min after tetanization in the presence of LY382884 was $45 \pm 10\%$ ($n = 3$), compared with $47 \pm 6\%$ in interleaved control experiments ($n = 6$; Fig. 3c). Furthermore, NMDA receptor-dependent LTP at CA1 synapses was also insensitive to LY382884 ($n = 7$; data not shown).

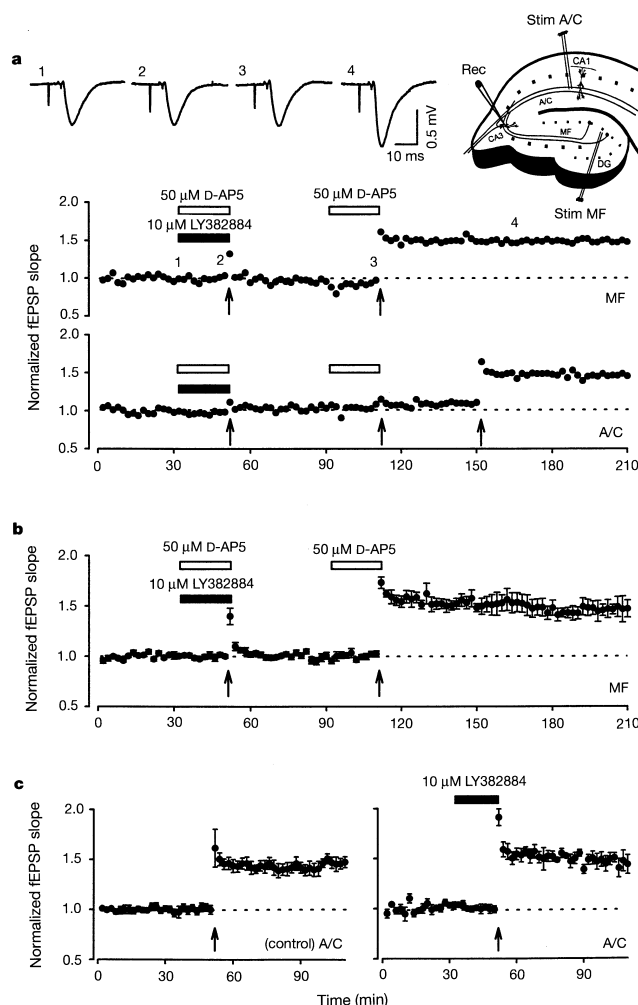


Figure 3 LY382884 specifically blocks the induction of mossy fibre LTP. **a**, A single example to show reversible block of the induction of mossy fibre (MF) LTP by LY382884. Note that the tetani were delivered in the presence of the NMDA receptor antagonist D-AP5, except for the third tetanus delivered to the associational/commissural (A/C) fibres. The traces in this and subsequent figures are averages of four successive records obtained at the times indicated on the graphs by the corresponding numbers. Inset, positions of the stimulating and recording electrodes. **b**, Pooled data for seven experiments showing reversible block of mossy-fibre-induced LTP. In this and subsequent graphs, data points are the average of four successive measurements and the time of each tetanus (100 Hz, 1 s, test intensity) is indicated by an arrow. **c**, The pooled data shows NMDA receptor-dependent LTP in the CA3 region of the hippocampus, evoked by tetanic stimulation of A/C fibres, under control conditions ($n = 6$) and in the presence of LY382884 ($n = 3$).

It was possible that LY382884 blocked the induction of LTP by an action that was independent of its ability to antagonize kainate receptors. Mossy fibre LTP involves the activation of the cAMP-protein kinase A (PKA) pathway and can be induced by the stimulation of adenylyl cyclase by forskolin¹⁹. Figure 4a shows input-specific, long-lasting (>4 h), PKA-dependent mossy fibre LTP induced by our standard tetanus protocol. LY382884 had no effect on forskolin-induced mossy fibre LTP ($n = 6$) when compared with interleaved controls ($n = 6$; Fig. 4b). Thus, its ability to inhibit LTP is unlikely to be due to an action on signal transduction or expression mechanisms involved in mossy fibre LTP.

The finding that kainate receptors are involved in the induction of mossy fibre LTP is unexpected, as two other antagonists, kynurenate and CNQX, have been reported not to block mossy-fibre-induced LTP^{20,21}. We confirmed that kynurenate could cause substantial antagonism of AMPA receptor-mediated synaptic transmission without preventing the induction of LTP, as determined following washout of kynurenate (this result was observed when 3 mM was applied for 20 min, our standard perfusion time to achieve a steady-state concentration; Fig. 5a). We verified that AMPA receptors are not required for induction of mossy fibre LTP using GYKI53655 (30 μ M, 20 min), a more selective AMPA receptor antagonist, which

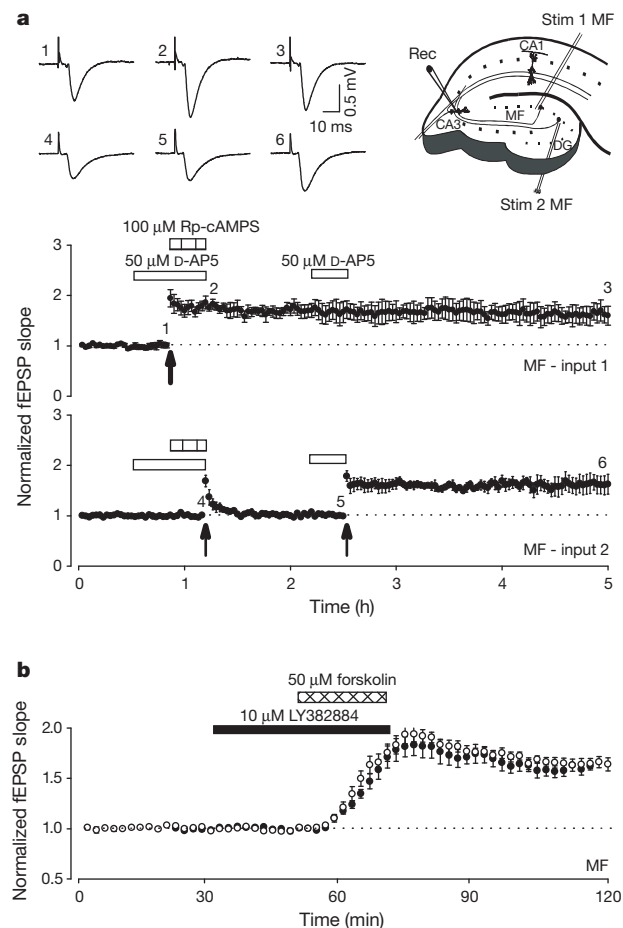


Figure 4 LY382884 does not affect forskolin-induced LTP. **a**, Pooled data ($n = 3$) to show that a single tetanus, delivered in the presence of D-AP5, induced input-specific, PKA-dependent LTP that lasted at least 4 h. Note that Rp-cAMPS (Rp-adenosine 3,5-cyclic monophosphothioate triethylamine) was applied immediately after tetanization of input 1 and washout was commenced immediately after tetanization of input 2. Inset, electrode arrangements. **b**, Pooled data to show that LY382884 does not affect LTP induced by application of forskolin. Data are from six slices treated with LY382884 (open symbols) and six interleaved controls (filled circles). Forskolin produced no potentiation of the associational/commissural input (data not shown).

reversibly blocked mossy fibre synaptic transmission. Three hours after the start of washout of GYKI53655, the non-tetanized input had recovered to baseline values ($+5 \pm 2\%$), whereas the tetanized input exhibited stable LTP (of $89 \pm 25\%$; $P < 0.05$; $n = 5$).

LTP was, however, fully blocked by 10 mM kynurenat (n = 5; Fig. 5b). We also found that 10 μ M CNQX, a concentration that antagonizes kainate-induced currents mediated by GluR5 in CA3 neurons¹⁶, blocked the induction of mossy fibre LTP (n = 4; Fig. 5c). In these two sets of experiments, we also examined the effects of LY382884 alone and of D(-)-2-amino-5-phosphonopentanoate (D-AP5) alone on mossy fibre LTP; in all cases, LY382884 blocked LTP whereas LTP was readily obtained in the presence of D-AP5 (the mean potentiation 60 min after tetanization in the presence of LY382884 alone was $-2 \pm 1\%$; 60 min after tetanization in the presence of D-AP5 alone it was $36 \pm 11\%$; n = 9). CNQX (K_i for displacement of GluR5 binding, $2.9 \pm 0.2 \mu$ M; n = 3) is about as effective as LY382884, whereas kynurenat is very weak as an antagonist at GluR5-expressing cells, with dose-dependent antagonism over the millimolar range. In three separate systems, kynurenat produced significantly ($P < 0.05$) more antagonism at 10 mM than at 3 mM: displacement of 100- μ M kainate binding in GluR5-expressing HEK293 cells was $98 \pm 1\%$ at 10 mM kynurenat and $60 \pm 1\%$ at 3 mM (n = 3); inhibition of 100 μ M kainate currents in DRG neurones was 89 ± 3 and $76 \pm 1\%$, respectively (n = 3); depression of synaptic responses mediated by kainate receptors in CA3 neurons was 84 ± 7 and $42 \pm 8\%$, respectively (n = 4). Thus, three structurally unrelated compounds antagonized both events mediated by kainate receptors containing GluR5 and mossy fibre LTP in the same rank order of potency: LY382884 = CNQX $\gg$ kynurenat.

LY382884 shows selectivity between homomeric GluR5 and GluR2 receptors¹⁰; it has been used in studies of global ischaemia¹⁰ and nociception^{22,23}. We have shown that LY382884 antagonizes responses mediated by kainate receptors at concentrations below those that affect synaptic processes mediated by AMPA or NMDA receptors. Like LY293558 (ref. 8) and LY294486 (ref. 9), LY382884 is highly selective for the GluR5 kainate receptor subunit. The earlier kainate receptor antagonists have been used to identify a role for GluR5 subunits in excitatory synaptic transmission in the hippocampus^{24,25} and amygdala²⁶. However, these compounds also antagonize AMPA receptors. Given its greater selectivity, LY382884 is a more useful antagonist with which to explore the functions of GluR5 in the brain.

Although our data indicate that the kainate receptor involved in the induction of LTP may have at least one GluR5 subunit, this does not preclude roles for other kainate receptor subunits in synaptic plasticity. Indeed, as the kainate receptor-mediated component of excitatory synaptic transmission at mossy fibre synapses^{16,17} is both antagonized by GluR5 antagonists²⁴ and absent in GluR6 knockout mice²⁷, both subunits may contribute to native kainate receptors in these neurons. The sensitivity of a GluR5/GluR6 heteromer to LY382884 is consistent with this possibility. Our findings do not indicate whether mossy fibre LTP is induced pre- or postsynaptically^{28,29}, as GluR5-containing kainate receptors are located at both sites. However, it may be possible to develop antagonists that are selective for pre- or postsynaptic kainate receptors at mossy fibres to address this question. Other receptors, such as mGlu receptors, may activate separate transduction processes that are also required for mossy fibre LTP^{20,29,30}. Hitherto, LTP has been viewed as usually being induced by the synaptic activation of NMDA receptors and expressed as an enhancement of synaptic transmission mediated by AMPA, and under certain circumstances NMDA receptors³. At synapses that exhibit LTP independently of NMDA receptors, it has been assumed that ionotropic glutamate receptors are not required for the induction of LTP^{28,29}. We have shown that kainate receptors are involved in the induction of LTP at a synapse where NMDA receptors do not have a role. It will be

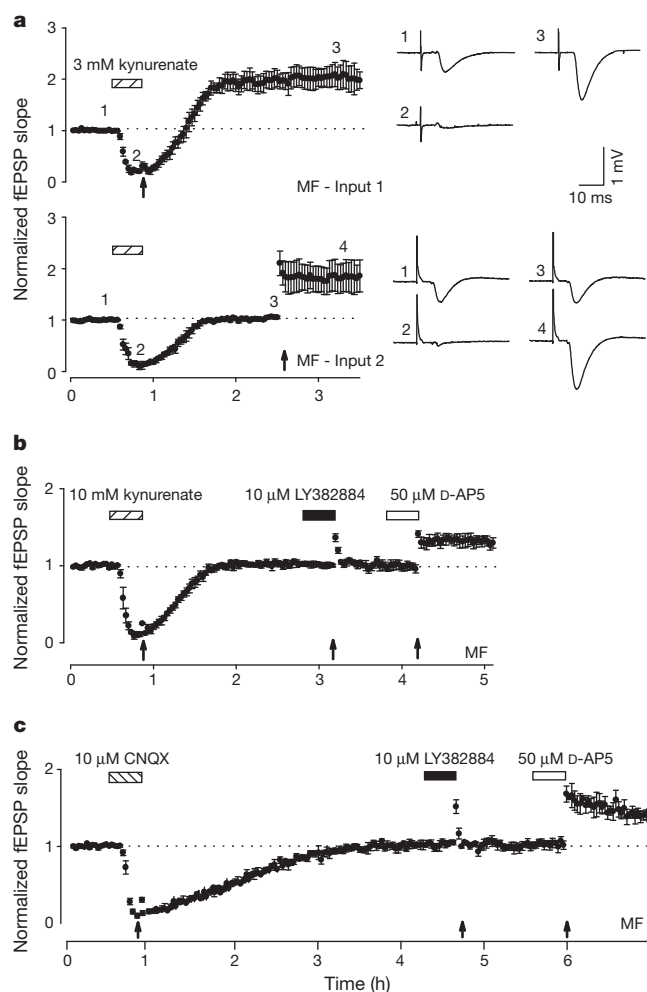


Figure 5 Kynurenat and CNQX block the induction of mossy fibre LTP. **a**, Pooled data ($n = 4$) to show that 3 mM kynurenat blocks AMPA receptor-mediated synaptic transmission but not mossy fibre LTP. A tetanus was delivered to input 1 during blockade of synaptic transmission and to input 2 following recovery from the effects of kynurenat. The traces are from a single experiment. **b**, Pooled data ($n = 5$) to show that 10 mM kynurenat, like 10 μ M LY382884, blocks the induction of mossy fibre LTP. **c**, Pooled data ($n = 4$) to show that 10 μ M CNQX, like 10 μ M LY382884, blocks the induction of mossy fibre LTP.

interesting to determine whether this principle extends to other synapses where NMDA receptor-independent LTP has been found and whether, by analogy to the NMDA receptor, the biophysical properties of kainate receptors^{5–7} endow properties of functional significance to this form of LTP.

Methods

Ligand-binding studies

These experiments were performed using cell membranes prepared from frozen HEK293 cells expressing recombinant AMPA or kainate receptor subunits and using [³H]-AMPA and [³H]-kainate, respectively, as described⁷.

Construction of the heteromer

A GluR6(Q) stable HEK293 cell line was transfected with GluR5_{2b}(R) in the vector pCEP4, using Lipofectamine 2000. After three weeks of selection with 250 μ g ml⁻¹ hygromycin, the cells reached 30% confluency and were split for electrophysiology.

Electrophysiology using isolated cells

Whole-cell voltage clamp recordings were made using extracellular solutions comprising (in mM) NaCl (138), CaCl₂ (5), KCl (5), MgCl₂ (1), HEPES (10) and glucose (10); pH 7.4.

For HEK293 cells and hippocampal neurons, intracellular solutions comprised (in mM) CsCl (140), MgCl₂ (1), diTris creatine phosphate (14), HEPES (10), BAPTA (15) and creatine phosphokinase (50 units ml⁻¹); pH 7.15. For DRG neurons⁸, intracellular solutions comprised (in mM) CsMeSO₄ (125), CsCl (15), CsBAPTA (5), HEPES (10), CaCl₂ (0.5) MgCl₂ (3) and MgATP (2); pH 7.2. Experiments were performed at room temperature (20–22 °C). Drugs were applied by bath perfusion and exchange of solutions under these conditions took about 5 s. Voltage ramps were conducted between –100 and +100 mV in 1 s. Experiments using recombinant receptors and DRG neurons were performed after incubation of cells with 250 µg ml⁻¹ concanavalin A for 10 min to remove receptor desensitization. Hippocampal pyramidal neurons were cultured from E17 Sprague Dawley rat fetuses. AMPA receptor-mediated currents were obtained from cells (6–12 days *in vitro*) in the presence of tetrodotoxin (1 µM); NMDA receptor-mediated currents were recorded from cells (10–12 days *in vitro*) in the presence of glycine (10 µM) but without added magnesium. Curve fitting to the data points was based upon the equation $y = 100(D^n/(D^n + EC_{50}^n))$ using a slope fixed to a value of 1 and where D is the drug concentration. For antagonists, $EC_{50} = 1/IC_{50}$. IC_{50} values were estimated from data obtained from at least four separate cells.

Electrophysiology in slice

Experiments were performed on transverse rat hippocampal slices (400 µm) maintained in medium comprising (in mM) NaCl (124), KCl (3), NaHCO₃ (26), NaH₂PO₄ (1.25), CaCl₂ (2), MgSO₄ (1) and D-glucose (10) (bubbled with O₂/CO₂: 95/5%). Extracellular fEPSPs were recorded in areas CA1 and CA3 using glass microelectrodes (2–4 MΩ) containing 4 M NaCl, as described¹⁴. Intracellular recordings were obtained using sharp glass microelectrodes (40–80 MΩ) filled with KMeSO₄ (2M)¹³. Whole-cell patch-clamp recordings were obtained blind using glass microelectrodes (5–7 MΩ; seal resistance ~10 GΩ) filled with a solution comprising (in mM) CsMeSO₃ (120), NaCl (1), MgCl₂ (1), Mg-ATP (4), BAPTA (10), N-(2,6-dimethyl-phenylcarbamoylmethyl)-triethylammonium bromide (QX-314) (5) and HEPES (5), adjusted to pH 7.3, as described¹⁶. Data are presented as mean ± s.e.m. throughout.

Received 10 September; accepted 4 October 1999.

- Hebb, D. O. *The Organization of Behaviour* (John Wiley, New York, 1949).
- Watkins, J. C. & Evans, R. H. Excitatory amino acid transmitters. *Annu. Rev. Pharmacol. Toxicol.* **21**, 165–204 (1981).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Bear, M. F. & Abraham, W. C. Long-term depression in hippocampus. *Annu. Rev. Neurosci.* **19**, 437–462 (1996).
- Bettler, B. & Mülle, C. AMPA and kainate receptors. *Neuropharmacology* **34**, 123–139 (1995).
- Bleakman, D. & Lodge, D. Neuropharmacology of AMPA and kainate receptors. *Neuropharmacology* **37**, 1187–1204 (1998).
- Chittajallu, R., Braithwaite S. P., Clarke, V. R. J. & Henley, J. M. Kainate receptors: subunits, synaptic localization and function. *Trends Pharmacol. Sci.* **20**, 26–35 (1999).
- Bleakman, D. *et al.* Pharmacological discrimination of GluR5 and GluR6 kainate receptor subtypes by (3S,4aR,6R,8aR)-6-[2-(1,2,4,5-tetrahydro-5-yl)ethyl]decahydroisoquinoline-3-carboxylic acid. *Mol. Pharmacol.* **49**, 581–585 (1996).
- Clarke, V. R. J. *et al.* A hippocampal GluR5 kainate receptor regulating inhibitory synaptic transmission. *Nature* **389**, 599–603 (1997).
- O'Neill, M. J. *et al.* Decahydroisoquinolines: novel competitive AMPA/kainate antagonists with neuroprotective effects in global cerebral ischaemia. *Neuropharmacology* **37**, 1211–1222 (1998).
- Huettnet, J. E. Glutamate receptor channels in rat DRG neurons: Activation by kainate and quisqualate and blockade of desensitization by Con A. *Neuron* **5**, 255–266 (1990).
- Partin, K. M. *et al.* Selective modulation of desensitization at AMPA versus kainate receptors by cyclothiazide and concanavalin A. *Neuron* **11**, 1069–1082 (1993).
- Davies, S. N. & Collingridge, G. L. Role of excitatory amino acid receptors in synaptic transmission in area CA1 of rat hippocampus. *Proc. R. Soc. Lond. B* **236**, 373–384 (1989).
- Vignes, M. *et al.* The GluR5 subtype of kainate receptor regulates excitatory synaptic transmission in areas CA1 and CA3 of the rat hippocampus. *Neuropharmacology* **37**, 1269–1277 (1998).
- Monaghan, D. T. & Cotman, C. W. The distribution of [³H]kainic acid binding sites in rat CNS as determined by autoradiography. *Brain Res.* **252**, 91–100 (1982).
- Vignes, M. & Collingridge, G. L. The synaptic activation of kainate receptors. *Nature* **388**, 179–182 (1997).
- Castillo, P. E., Malenka, R. C. & Nicoll, R. A. Kainate receptors mediate a slow postsynaptic current in hippocampal CA3 neurons. *Nature* **388**, 182–186 (1997).
- Harris, E. W. & Cotman, C. W. Long-term potentiation of guinea pig mossy fiber responses is not blocked by N-methyl-D-aspartate antagonists. *Neurosci. Lett.* **70**, 132–137 (1986).
- Weisskopf, M. G., Castillo, P. E., Zalutsky, R. A. & Nicoll, R. A. Mediation of hippocampal mossy fiber long-term potentiation by cyclic AMP. *Science* **265**, 1878–1882.
- Ito, I. & Sugiyama, H. Roles of glutamate receptors in long-term potentiation at hippocampal mossy fiber synapses. *NeuroReport* **2**, 333–336 (1991).
- Castillo, P. E., Weisskopf, M. G. & Nicoll, R. A. The role of Ca²⁺ channels in hippocampal mossy fiber synaptic transmission and long-term potentiation. *Neuron* **12**, 261–269 (1994).
- Simmons, R. M. A. *et al.* Kainate GluR5 receptor subtype mediates the nociceptive response to formalin in the rat. *Neuropharmacology* **37**, 25–36 (1998).
- Procter, M. J. *et al.* Actions of kainate and AMPA selective glutamate receptor ligands on nociceptive processing in the spinal cord. *Neuropharmacology* **37**, 1287–1297 (1998).
- Vignes, M., Bleakman, D., Lodge, D. & Collingridge, G. L. The synaptic activation of the GluR5 subtype of kainate receptor in area CA3 of the rat hippocampus. *Neuropharmacology* **36**, 1477–1481 (1997).
- Cossart, R. *et al.* GluR5 kainate receptor activation in interneurons increases tonic inhibition of pyramidal cells. *Nature Neurosci.* **1**, 470–478 (1998).

- Li, H. & Rogawski, M. A. GluR5 kainate receptor mediated synaptic transmission in rat basolateral amygdala *in vitro*. *Neuropharmacology* **37**, 1279–1286 (1998).
- Mulle, C. *et al.* Altered synaptic physiology and reduced susceptibility to kainate-induced seizures in GluR6-deficient mice. *Nature* **392**, 601–605 (1998).
- Nicoll, R. A. & Malenka, R. C. Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature* **377**, 115–118 (1995).
- Yeckel, M. F., Kapur, A. & Johnston, D. Multiple forms of LTP in hippocampal CA3 neurons use a common postsynaptic mechanism. *Nature Neurosci.* **2**, 625–633 (1999).
- Bashir, Z. I. *et al.* Induction of LTP in the hippocampus needs synaptic activation of glutamate metabotropic receptors. *Nature* **363**, 347–350 (1993).

Acknowledgements

We thank B. Anderson for providing the data acquisition software (<http://www.ltp-program.com>). This work was supported by the MRC. I.S. is a postdoc of the FWO-Vlaanderen, Belgium.

Correspondence and requests for materials should be addressed to Z.A.B.

Activation of the epithelial Na⁺ channel (ENaC) requires CFTR Cl⁻ channel function

M. M. Reddy, M. J. Light & P. M. Quinton

Department of Pediatrics-0831, University of California, San Diego, School of Medicine, La Jolla, California 92093-0831, USA

It is increasingly being recognized that cells coordinate the activity of separate ion channels that allow electrolytes into the cell. However, a perplexing problem in channel regulation has arisen in the fatal genetic disease cystic fibrosis, which results from the loss of a specific Cl⁻ channel (the CFTR channel) in epithelial cell membranes¹. Although this defect clearly inhibits the absorption of Na⁺ in sweat glands^{2,3}, it is widely accepted that Na⁺ absorption is abnormally elevated in defective airways in cystic fibrosis^{4,5}. The only frequently cited explanation for this hypertransport is that the activity of an epithelial Na⁺ channel

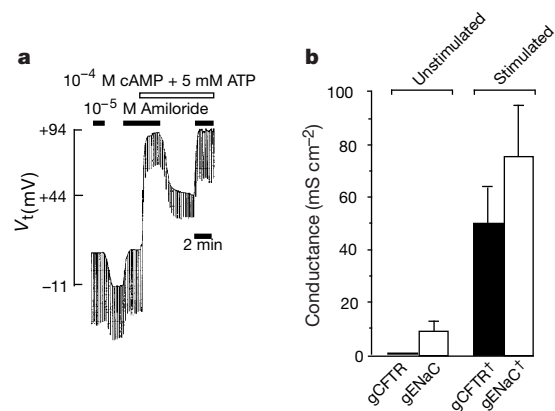


Figure 1 Effect of cAMP and ATP activation of CFTR Cl⁻ conductance (gCFTR) on amiloride-sensitive epithelial Na⁺ conductance (gENaC). **a**, Diffusion potentials and constant current–pulse potentials across the apical membrane. Without stimulation, CFTR is completely deactivated, but a small stimulation-insensitive, amiloride-sensitive Na⁺ diffusion potential and gENaC remain. However, stimulation results in a significant increase in both gCFTR and gENaC (see text). **b**, Mean conductances of gCFTR (filled columns) and gENaC (open columns) measured under baseline conditions (Unstimulated) and after gCFTR when stimulated with 0.1 mM cAMP plus 5 mM ATP (Stimulated) ($P < 0.01$, $n = 15$).

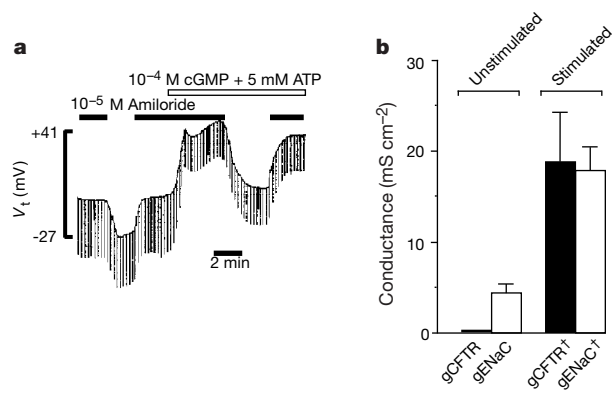


Figure 2 Effect of stimulating gCFTR with cGMP and ATP on gENaC. **a**, The apical membrane potential shows that both gCFTR and gENaC increase when stimulated with 0.1 mM cGMP plus ATP. **b**, Mean specific conductances for gCFTR and gENaC before (Unstimulated) and after (Stimulated) stimulation with cGMP ($P < 0.05$, $n = 3$).

(ENaC) is inversely related to the activity of the CFTR Cl^- channel^{5–7}. However, we report here that, in freshly isolated normal sweat ducts, ENaC activity is dependent on, and increases with, CFTR activity. Surprisingly, we also find that the primary defect in Cl^- permeability in cystic fibrosis⁸ is accompanied secondarily by a Na^+ conductance in this tissue that cannot be activated. Thus, reduced salt absorption in cystic fibrosis is due not only to poor Cl^- conductance but also to poor Na^+ conductance.

It is fundamental in net electroconductive salt transport that cations and anions move separately but in parallel to maintain electroneutrality. In the transport of an ionized salt such as NaCl, therefore, impeding the movement of either Na^+ or Cl^- should impede the movement of the other. The inherent loss of Cl^- conductance in cell membranes of sweat ducts from cystic fibrosis patients^{3,4,8} has been assumed to impede the uptake of Na^+ through purely electrochemical effects. Consequently, the absorption of both Na^+ and Cl^- in sweat ducts is significantly reduced, and salt appears in the sweat at diagnostic concentrations of up to five times higher than normal.

However, despite the loss of Cl^- permeability, several observations, such as amiloride inhibition of transepithelial electrical potentials⁹, larger short-circuit currents¹⁰, and increased Na^+ flux have been interpreted as reflecting increased Na^+ absorption across airway epithelia in cystic fibrosis patients^{4,5,11}. Recent studies on heterologous systems have offered a potential explanation for these counterintuitive results as the expression and/or activity of the CFTR channel appeared to inhibit ENaC activity in these systems^{5,6,12}. Accordingly, it has been widely accepted that the loss of CFTR function causes increased ENaC activity, resulting in hyperabsorption of Na^+ (refs 3–5). The human sweat duct provides a unique opportunity to test this hypothesis in a native tissue. The normal duct is composed of a syncytium of a single set of homogeneously similar cells dedicated to salt absorption, where CFTR and ENaC are the only active conductances seen in the apical membrane^{13–15}, so other ion conductances are not likely to confound interpretation. In contrast to normal ducts, functional CFTR is almost completely absent in cystic fibrosis sweat-duct cells^{2,13}, so the native properties of ENaC can be observed in the virtual absence of Cl^- conductance.

We tested the physiological interrelations between ENaC and CFTR by using three different means of activating CFTR in the apical membrane of normal ducts. First, endogenous cyclic nucleotides and ATP were depleted from the cell through selectively permeabilized basolateral membranes, deactivating CFTR. We

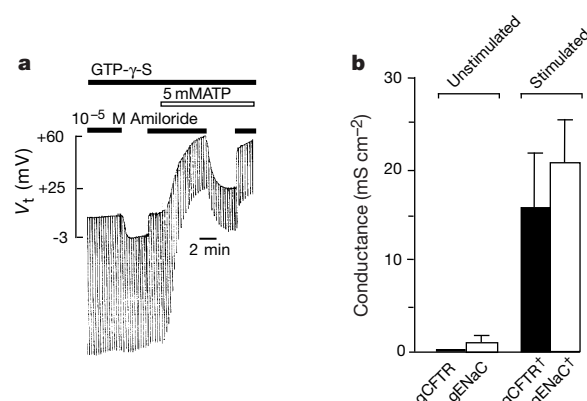


Figure 3 Effect of stimulating gCFTR with GTP- γ -S and ATP on gENaC. **a**, Record in which activating CFTR by G proteins also activates gENaC. **b**, Mean gENaC and gCFTR before (Unstimulated) and after (Stimulated) activation with 100 μM GTP- γ -S plus 5 mM ATP ($P < 0.02$, $n = 4$).

then reactivated CFTR by adding back either: cyclic AMP to stimulate protein kinase A¹⁵; cyclic GMP to stimulate protein kinase G¹⁶; or GTP- γ -S¹⁶ to activate heterotrimeric guanine-nucleotide-binding proteins (G proteins). To determine the effect of CFTR activity on ENaC, we measured the Na^+ conductance (gENaC) before and after activating the CFTR Cl^- conductance (gCFTR). We recorded the changes in transepithelial electrodiffusion potentials and conductances in response to amiloride, which is a selective ENaC inhibitor, as a direct measure of gENaC. Activating gCFTR by any of the three agonists results in simultaneous increases in gENaC that were approximately equal to the corresponding increases in gCFTR (Figs 1–3). When gCFTR was activated by cAMP, the average gENaC increased by 67 mS cm^{-2} and gCFTR increased by 51 mS cm^{-2} (Fig. 1). When CFTR was activated with cGMP (Fig. 2) or G-protein agonists, gENaC increased by about 13 and 20 mS cm^{-2} , respectively (Fig. 3) as gCFTR increased by 20 and 16 mS cm^{-2} , respectively. Hence, activating gCFTR was coupled with roughly equal increases in gENaC. Conversely, deactivating gCFTR (by removing the agonists) resulted in marked reductions in gENaC (back to baseline values). These results show that gCFTR and gENaC are activated in parallel and that, in this native tissue, gENaC is not inhibited when gCFTR increases, as proposed in other systems^{5,6,10,11}.

Because CFTR activation requires protein kinase A (PKA) phosphorylation, we investigated whether the coupling of CFTR and gENaC might simply be due to simultaneous phosphorylation activation^{12,17}. We were able to test whether ENaC activates by phosphorylation, independently of CFTR, because CFTR activation also requires a physiological concentration of ATP (~ 5 mM), whereas PKA phosphorylation activity requires much lower concentrations of ATP (K_m , 3.1 μM)¹⁸. When we added 100 μM ATP plus 100 μM cAMP to the duct cell cytoplasm to activate PKA without activating CFTR, CFTR did not activate and gENaC did not increase. Before and after activating PKA, gENaC was 6 ± 1 mS cm^{-2} and 5 ± 1 mS cm^{-2} , respectively, in normal ducts ($n = 4$, $p > 0.4$).

To ensure that phosphorylation activation of ENaC did not fail simply as a result of the lower ATP concentration, we repeated these experiments in the presence of 5 mM ATP with ducts from cystic fibrosis patients that inherently lack functional CFTR. We compared gENaC in normal and cystic fibrosis ducts before and after stimulation with cAMP and ATP (as in Fig. 1) at concentrations that are optimal for activating CFTR in non-cystic fibrosis ducts^{14,15}. We found that gENaC in cystic fibrosis ducts did not change after stimulation (5 ± 1 mS cm^{-2} before and 4 ± 1 mS cm^{-2} after stim-

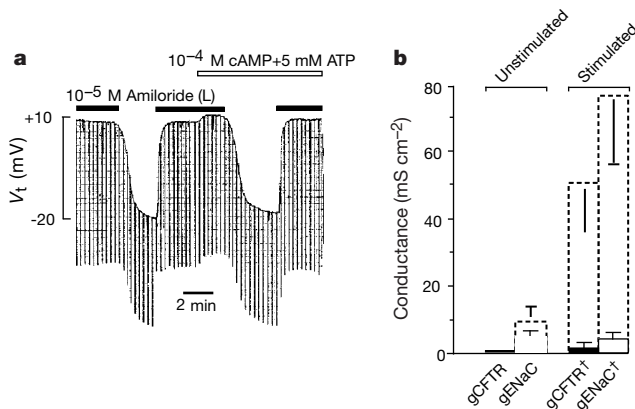


Figure 4 Failure of stimulation to increase gENaC in the absence of functional CFTR. **a**, Record of the apical membrane potential for a cystic fibrosis duct. **b**, Mean gCFTR and gENaC in cystic fibrosis ducts before (Unstimulated) and after (Stimulated) stimulation with cAMP plus ATP. For comparison, the corresponding means from Fig. 1b are shown as dashed columns. gENaC in unstimulated normal ducts was about twice that in cystic fibrosis ducts. ($n = 7$ cystic fibrosis ducts, $P > 0.08$). After stimulation, gENaC increased tenfold in normal ducts, but was unaffected in cystic fibrosis ducts ($P < 0.01$).

ulation). This lack of response in cystic fibrosis ducts contrasted dramatically with the large increases in gENaC and gCFTR observed in non-cystic fibrosis ducts (Fig. 1). These results (Fig. 4) show either that PKA does not control the activation of ENaC directly, or that PKA phosphorylation of ENaC is dependent on activated CFTR. These results are surprising, as it has long been assumed that the failure to absorb NaCl in cystic fibrosis ducts was due solely to the loss of Cl^- conductance. Our results indicate that the failure also includes an almost equivalent loss of Na^+ conductance, owing to its dependence on CFTR activation.

From the point of view of satisfying electroneutrality, it is consistent that CFTR and ENaC channels should open together when NaCl (Na^+ and Cl^-) is absorbed electroconductively. By the same rationale, reciprocal activities of CFTR and ENaC channels seem unfavourable, if not contra-indicated, for efficient salt absorption. However, it is not clear why CFTR and ENaC should function reciprocally in heterologous cells and airway epithelia^{4,6}. It is possible that variant forms of ENaC or different regulatory components operate in different systems. For example, immortalized distal tubule A6-C1 cells show proportional but temporally separated changes in Cl^- and Na^+ conductances in response to vasopressin¹⁹, whereas cultures of M-1 mouse cortical collecting duct cells²⁰ and dexamethasone-treated rat colonocytes²¹ exhibit reciprocity.

However, heterologous systems with overexpressed proteins may adversely affect several cell functions. For example, in CFPAC-1 cells, hyperexpressing CFTR depressed both outward-rectifying Cl^- currents and inward-rectifying K^+ currents²². Our finding that ENaC and CFTR activation are closely coupled is indicative of intermolecular interactions. The carboxy-terminal motif interacting with PDZ1 or PDZ2 of ERM-binding phosphoprotein 50 or other related proteins may be important for CFTR function^{23,24}. If any of the putative components that facilitate interactions between CFTR and ENaC are not expressed or are aligned differently, then we might expect different functions in different systems. It remains to be determined whether diverse alignments between CFTR and ENaC can explain the paradox of cystic fibrosis airway epithelial cells^{4,5,25}, or whether they act consistently in parallel in electroconductive absorption, requiring coupled CFTR and ENaC activities. A native airway epithelium both secretes and absorbs NaCl ^{8,26}. The proportional activities of ENaC and CFTR seen in the sweat duct might therefore be present in airway absorptive cells but overshadowed by altered functions in other cells; that is, cAMP-

mediated secretion is virtually absent in cystic fibrosis exocrine epithelia^{3,4,27}. A loss of secretory transport, even in the presence of simultaneously diminished absorptive transport, could therefore result in an apparently larger Na^+ absorption in cystic fibrosis airway tissue¹¹.

Our results show that ENaC function depends critically on the state of CFTR in a purely salt-absorbing epithelium and that ENaC cannot be activated in cystic fibrosis ducts where CFTR is defective or absent. Poor CFTR function is therefore not characteristically coupled to abnormally elevated Na^+ absorption and increased gENaC in this disease. Failure to absorb NaCl in cystic fibrosis sweat ducts is therefore due not only to the electrochemical effects of a primary loss of gCFTR but also to a secondary inability to activate gENaC. □

Methods

Sweat ducts were dissected and microperfused as described^{14,15}. Duct segments were cannulated with a double-barrelled micropipette so that luminal perfusate could be changed and the transepithelial potential monitored through one barrel while constant current pulses were applied through the other. Conductance was calculated from the cable equation^{13–15,28}. We selectively permeabilized the basolateral membrane with 1,000–5,000 U ml^{-1} α -toxin (from *Staphylococcus aureus*) so that cytoplasmic concentrations of cAMP, cGMP, GTP- γ -S and ATP and electrolytes could be controlled. The conductance of the remaining apical membrane consisted of gENaC, gCFTR and a small non-selective leak^{14,15}. Ion gradients were established for both Na^+ and Cl^- across the apical membrane by perfusing the cytoplasmic bath with NaCl-free solution containing (in mM): K^+ , 145; gluconate, 140; PO_4^{3-} , 3.5; MgSO_4 , 1.2; Ca^{2+} , 0.26; EGTA, 2.0; pH 6.8. We simultaneously perfused the lumen with Ringer's solution containing (in mM): NaCl, 150; K^+ , 5; PO_4^{3-} , 3.5; MgSO_4 , 1.2; Ca^{2+} , 1; pH 7.4. The effect of permeabilizing the basolateral membrane to deplete intracellular cAMP and ATP to deactivate CFTR is shown in Fig. 1. Amiloride was always present in the luminal perfusate except when gENaC and Na^+ diffusion potentials were measured. gCFTR and gENaC were calculated using the following formulae: $G_i = g\text{ENaC} + g\text{X}$, and $G_i \uparrow = g\text{CFTR} + g\text{ENaC} \uparrow + g\text{X}$, where G_i and $G_i \uparrow$ are the total transepithelial conductances before and after activating CFTR with cAMP, cGMP or GTP- γ -S plus ATP; gX is a non-specific leak conductance measured after blocking gENaC (with amiloride) and gCFTR (by removing cAMP and ATP); gENaC and gENaC $\uparrow$ represent ENaC conductances before and after activating CFTR. Probabilities less than 0.05 (determined by Student *t*-test for paired and unpaired means as appropriate) were taken as significant.

Received 22 April; accepted 29 September 1999.

- Sheppard, D. N. & Welsh, M. J. Structure and function of the CFTR chloride channel. *Physiol. Rev.* **79**, S23–S45 (1999).
- Quinton, P. M. Chloride impermeability in cystic fibrosis. *Nature* **301**, 421–422 (1983).
- Quinton, P. M. Physiological basis of cystic fibrosis: A historic perspective. *Physiol. Rev.* **79**, S3–S22 (1999).
- Boucher, R. C. Human airway ion transport. Part I. *Am. J. Respir. Crit. Care Med.* **150**, 271–281 (1994).
- Schwiebert, E. M., Benos, D. J., Egan, M. E., Stutts, M. J. & Guggino, W. B. CFTR is a conductance regulator as well as a chloride channel. *Physiol. Rev.* **79**, S145–S166 (1999).
- Stutts, M. J. *et al.* CFTR as a cAMP-dependent regulator of sodium channels. *Science* **269**, 847–850 (1995).
- Kunzelmann, K., Kiser, G. L., Schreiber, R. & Riordan, J. R. Inhibition of epithelial Na^+ currents by intracellular domains of the cystic fibrosis transmembrane conductance regulator. *FEBS Lett.* **400**, 341–344 (1997).
- Pilewski, J. M. & Frizzell, R. A. Role of CFTR in airway disease. *Physiol. Rev.* **79**, S215–S255 (1999).
- Knowles, M. R. *et al.* Abnormal ion permeation through cystic fibrosis respiratory epithelium. *Science* **221**, 1067–1070 (1983).
- Boucher, R. C., Cotton, C. U., Gatzky, J. T., Knowles, M. R. & Yankaskas, J. R. Evidence for reduced Cl^- and increased Na^+ permeability in cystic fibrosis human primary cell cultures. *J. Physiol. (Paris)* **405**, 77–103 (1988).
- Boucher, R. C., Stutts, M. J., Knowles, M. R., Cantley, L. & Gatzky, J. T. Na^+ transport in cystic fibrosis respiratory epithelial: abnormal basal rate and response to adenylate cyclase activation. *J. Clin. Invest.* **78**, 1245–1252 (1986).
- Stutts, M. J., Rossier, B. C. & Boucher, R. C. Cystic fibrosis transmembrane conductance regulator inverts protein kinase A-mediated regulation of epithelial sodium channel single channel kinetics. *J. Biol. Chem.* **272**, 14037–14040 (1997).
- Quinton, P. M. Missing Cl^- conductance in cystic fibrosis. *Am. J. Physiol.* **251**, C649–C652 (1986).
- Quinton, P. M. & Reddy, M. M. Control of CFTR Cl^- conductance by ATP levels through non-hydrolytic binding. *Nature* **360**, 79–81 (1992).
- Reddy, M. M. & Quinton, P. M. Rapid regulation of electrolyte absorption in sweat duct. *J. Membr. Biol.* **140**, 57–67 (1994).
- Reddy, M. M. & Quinton, P. M. G-Proteins activate CFTR- G_{Cl} in the native sweat duct. *Ped. Pulmonol.* **S14**, R46 (1998).
- Ismailov, I. I., McDuffie, J. H. & Benos, D. J. Protein kinase A phosphorylation and G-protein regulation of purified renal Na^+ channels in planar bilayer membranes. *J. Biol. Chem.* **269**, 10235–10241 (1994).
- Flockhart, D. A., Friest, W., Hoppe, J., Loncoln, T. M. & Corban, J. D. ATP analog specificity of cAMP dependent protein kinase, cGMP dependent protein kinase, and phosphorylase kinase. *Eur. J. Biochem.* **140**, 289–295 (1984).

19. Verrey, F. Antidiuretic hormone action in A6 cells: effect on apical Cl and Na conductances and synergism with aldosterone for NaCl reabsorption. *J. Membr. Biol.* **138**, 65–76 (1994).
20. Letz, B., Korbmayer, C. cAMP stimulates CFTR-like Cl⁻ channels and inhibits amiloride-sensitive Na⁺ channels in mouse CCD cells. *Am. J. Physiol.* **272**, C657–C666 (1997).
21. Ecke, D., Bleich, M. & Greger, R. The amiloride-inhibitable Na⁺ conductance of rat colonic crypt cells is suppressed by forskolin. *Pflügers Arch.* **431**, 984–986 (1996).
22. Panah, M. R. *et al.* Hyperexpression of recombinant CFTR in heterologous cells alters its physiological properties. *Am. J. Physiol. Cell* **274**, C310–C318 (1998).
23. Short, D. B. *et al.* An apical PDZ protein anchors the cystic fibrosis transmembrane conductance regulator to the cytoskeleton. *J. Biol. Chem.* **273**, 19797–19801 (1998).
24. Hall, R. A. *et al.* A C-terminal motif found in the β_2 -adrenergic receptor, P2Y₁ receptor and cystic fibrosis transmembrane conductance regulator determines binding to the Na⁺/H⁺ exchanger regulatory factor family of PDZ proteins. *Proc. Natl Acad. Sci. USA* **95**, 8496–8501 (1998).
25. Jiang, C., Finkbeiner, W. E., Widdicombe, J. H. & Miller, S. S. Fluid transport across cultures of human tracheal glands is altered in cystic fibrosis. *J. Physiol. (Lond.)* **501**, 637–647 (1997).
26. Welsh, M. J. Electrolyte transport by airway epithelia. *Physiol. Rev.* **67**, 1143–1184 (1987).
27. Sato, K. & Sato, F. Defective beta adrenergic response of cystic fibrosis sweat glands *in vivo* and *in vitro*. *J. Clin. Invest.* **73**, 1763–1771 (1984).
28. Greger, R. & Schlatter, F. Properties of the basolateral membrane of the cortical thick ascending limb of Henle's loop of rabbit kidney. *Eur. J. Physiol.* **396**, 325–334 (1983).

Acknowledgements

We thank K. Taylor for technical assistance. This work was funded by grants from the Nancy McCracken Endowment, the Cystic Fibrosis Foundation and the USPHS National Institutes of Health.

Correspondence and requests for materials should be addressed to P.Q.
(e-mail: pquinton@ucsd.edu).

Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand

Young-Yun Kong^{*†}, Ulrich Feige[‡], Ildiko Sarosi[§], Brad Bolon[§], Anna Tafuri^{*}, Sean Morony[§], Casey Capparelli[§], Ji Li^{||}, Robin Elliott^{||}, Susan McCabe^{||}, Thomas Wong[¶], Giuseppe Campagnuolo[‡], Erika Moran[#], Earl R. Bogoch[#], Gwyneth Van[§], Linh T. Nguyen^{*}, Pamela S. Ohashi^{*}, David L. Lacey[§], Eleanor Fish[¶], William J. Boyle^{||} & Josef M. Penninger^{*☆}

^{*} Amgen Institute, 620 University Avenue, Toronto, Ontario M5G 2C1, Canada

[‡] Department of Pharmacology, [§] Pathology, and ^{||} Cell Biology, Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320-1789, USA

[¶] Department of Medical Genetics & Microbiology, [#] St Michael's Hospital

[☆] Ontario Cancer Institute and the Departments of Medical Biophysics and Immunology, University of Toronto, Toronto, Ontario, Canada

[†] These authors contributed equally to this work

Bone remodelling and bone loss are controlled by a balance between the tumour necrosis factor family molecule osteoprotegerin ligand (OPGL) and its decoy receptor osteoprotegerin (OPG)^{1–3}. In addition, OPGL regulates lymph node organogenesis, lymphocyte development and interactions between T cells and dendritic cells in the immune system^{3–5}. The OPGL receptor, RANK, is expressed on chondrocytes, osteoclast precursors and mature osteoclasts^{4,6}. OPGL expression in T cells is induced by antigen receptor engagement⁷, which suggests that activated T cells may influence bone metabolism through OPGL and RANK. Here we report that activated T cells can directly trigger osteoclastogenesis through OPGL. Systemic activation of T cells *in vivo* leads to an OPGL-mediated increase in osteoclastogenesis and bone loss. In a T-cell-dependent model of rat adjuvant arthritis characterized by severe joint inflammation, bone and cartilage destruction and crippling, blocking of OPGL through osteoprotegerin treatment at the onset of disease prevents bone

and cartilage destruction but not inflammation. These results show that both systemic and local T-cell activation can lead to OPGL production and subsequent bone loss, and they provide a novel paradigm for T cells as regulators of bone physiology.

Remodelling of bone involves the synthesis of bone matrix by osteoblasts and bone resorption by osteoclasts⁸. Excessive osteoclast activity is observed in many osteopenic disorders characterized by increased bone resorption and crippling bone damage, including post-menopausal osteoporosis⁹, Paget's disease¹⁰ and lytic bone metastases¹¹. Local or generalized bone loss has also been reported in chronic infections (hepatitis, HIV)¹², leukaemias¹³, autoimmune and allergic diseases¹⁴, and rheumatoid arthritis¹⁵, suggesting that an activated immune system can affect bone physiology. Although cytokines, such as tumour necrosis factor (TNF) α , interleukin (IL)-1, IL-11 and IL-17¹⁶, that regulate immune functions have been implicated in the regulation of bone homeostasis, the molecular mechanism by which the immune system affects bone metabolism has never been established.

Consistent with previous reports^{7,17}, we observed that OPGL (also known as TRANCE, RANK-L or ODF) messenger RNA expression was upregulated in murine T cells following antigen receptor engagement (Fig. 1a). Specific inhibitor studies showed that induction of OPGL was dependent on protein kinase C (PKC), phosphoinositide 3-kinase (PI3K) and calcineurin-mediated signalling pathways (Fig. 1a). Expression of OPGL protein was detected on the surfaces of activated, but not resting, T cells (Fig. 1b). Activated T cells also secreted soluble OPGL into culture medium (Fig. 1c), and soluble OPGL was detected by enzyme-linked immunosorbent assay (ELISA) ($\sim 2.2 \pm 0.2$ ng ml⁻¹ of OPGL in the culture supernatant of CD4⁺ T cells activated for 4 days with α CD3 plus α CD28). To determine whether the secreted and/or surface forms of OPGL were functionally active, we assessed the ability of both molecules to induce osteoclastogenesis *in vitro*. Haematopoietic bone marrow precursors from wild-type mice were co-cultured with activated CD4⁺ T cells fixed with paraformaldehyde or with culture supernatants from activated T cells. Both membrane-bound and soluble OPGL supported osteoclast development *in vitro* (Fig. 1d). Osteoclastogenesis was blocked in both cases in the presence of the physiological decoy receptor osteoprotegerin (OPG) (Fig. 1d). As certain T-cell-derived cytokines can regulate OPGL expression in stromal cells, we neutralized many different cytokines including IL-1, TNF α , IL-6, interferon (INF) γ , IL-3, and granulocyte/macrophage colony-stimulating factor (GM-CSF), and none of these cytokines proved critical for *in vitro* osteoclastogenesis in this culture system, confirming that T-cell-derived OPGL is the crucial mediator. To exclude the possibility that another surface receptor on fixed T cells activates stromal cells for OPGL expression, we used spleen cells from newborn *opgl*^{-/-} mice as a source of haematopoietic precursors and fixed wild-type CD4⁺ T cells. In this culture, the only source of OPGL is membrane-bound OPGL on the T cells. Wild-type T cells induced the formation of tartrate-resistant acid phosphatase (TRAP)⁺ osteoclasts from *opgl*^{-/-} progenitor cells, and osteoclastogenesis was inhibited by addition of OPG (Fig. 1d). Osteoclastogenesis was confirmed by detection of the osteoclast-specific markers calcitonin receptor (CTR), RANK (Fig. 1e), cathepsin K (Fig. 1f) and β_3 -integrin (Fig. 1g). Thus, T-cell-triggered osteoclasts display a typical osteoclast phenotype (TRAP⁺, CTR⁺, cathepsin K⁺, β_3 -integrin⁺, RANK⁺). These results provide genetic evidence that activated T cells can directly trigger osteoclastogenesis through membrane-bound and soluble OPGL.

To evaluate whether T-cell activation and expression of OPGL affect bone physiology *in vivo*, we explored the phenotype of *ctla4*^{-/-} mice, in which T cells are spontaneously activated¹⁸. All *ctla4*^{-/-} mice displayed severe osteoporosis compared with heterozygote littermates (Fig. 2a–d). To further explore the effect of spontaneously activated T cells on bone metabolism, we adoptively transferred *ctla4*^{+/-} or *ctla4*^{-/-} bone marrow cells into lymphocyte-deficient

rag1^{-/-} mice¹⁹ and examined bone structure eight weeks after the cell graft. *rag1*^{-/-} mice have normal numbers of osteoclasts, normal bone morphology and normal bone mineral densities (data not shown). Transfer of *ctla4*^{-/-} cells into *rag1*^{-/-} mice caused a significant decrease in total and trabecular bone mineral densities compared with control *ctla4*^{-/-}*rag1*^{-/-} chimaeras and *rag1*^{-/-} mice (Table 1). Histologically, *ctla4*^{-/-}*rag1*^{-/-} chimaeric mice exhibited extended resorption of trabecular bone below the growth plates and epiphyses of the femur and tibia (Fig. 2e–h). In addition to

decreased mineral density, actual bone loss was observed by histomorphometry (Table 2). Whereas only a few TRAP⁺ osteoclasts were detectable in *ctla4*^{-/-}*rag1*^{-/-} mice, osteoclast numbers were significantly increased in *ctla4*^{-/-}*rag1*^{-/-} mice (Table 2, and Fig. 2i, j). We observed a similar loss in bone after transfer of purified *ctla4*^{-/-} T cells into *opgl*^{-/-} mice (data not shown), implying that activated T cells can directly cause bone loss *in vivo*. Daily injection of *ctla4*^{-/-} mice with the decoy receptor OPG, which blocks OPG action, increased bone

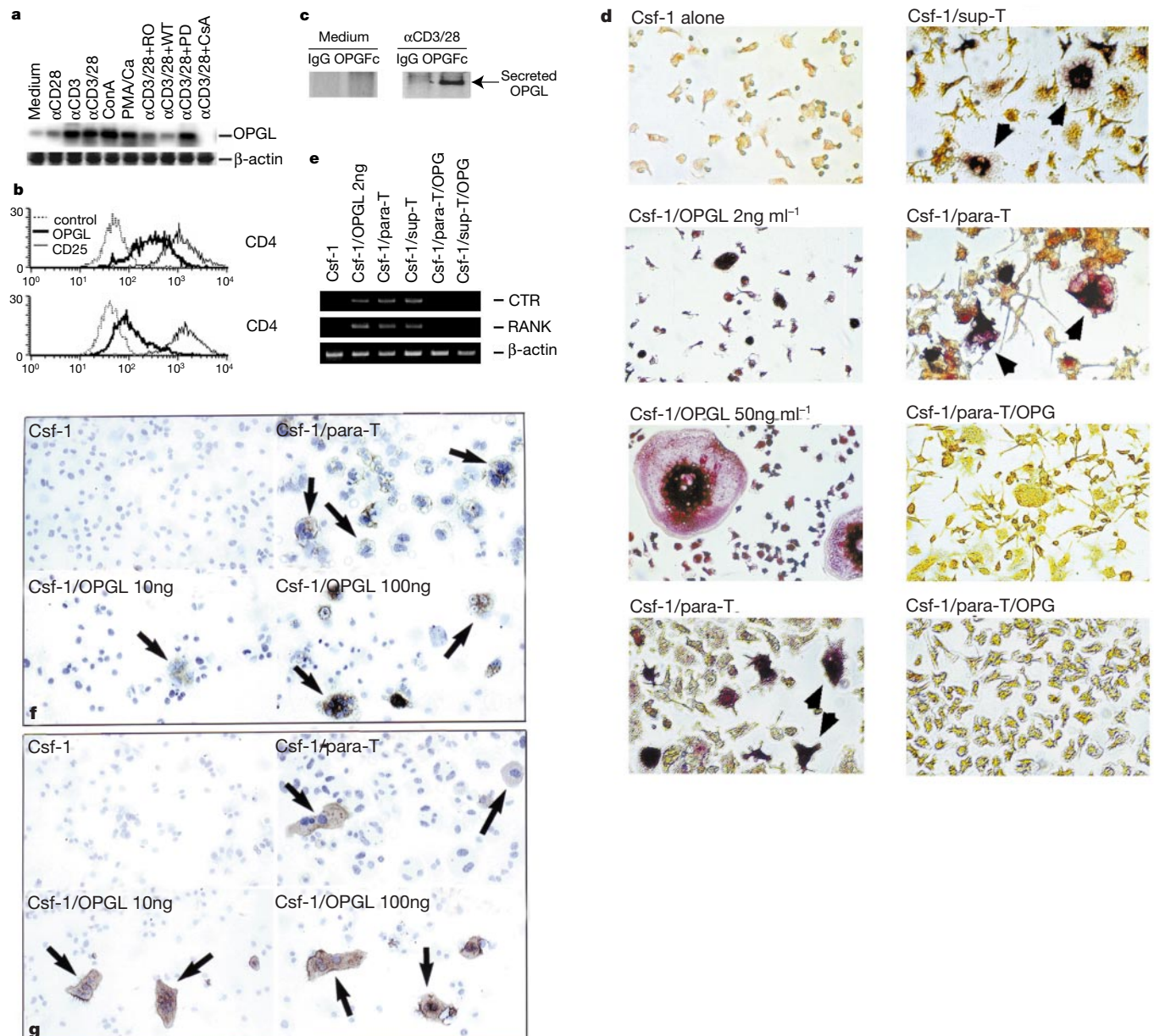


Figure 1 Activated T cells induce osteoclastogenesis through OPGL. **a**, Purified CD4⁺ lymph-node T cells were activated for 16 h with the indicated stimuli in the presence or absence of the signalling inhibitors RO, WT, PD and CsA (see Methods). OPGL and β -actin mRNA expression were detected by RT-PCR. **b**, Expression of OPGL on the surfaces of T cells stimulated with anti-CD3 ϵ plus anti-CD28 antibodies for 4 days. OPGL expression was detected by FACS. CD25 expression is shown to control for activation. **c**, Expression of secreted OPGL protein in supernatants of control T cells (left) or T cells stimulated with anti-CD3 plus anti-CD28 for 2 days (right). **d**, *In vitro* osteoclastogenesis. Bone marrow precursors from wild-type mice (top six panels) and newborn *opgl*^{-/-} mice (bottom two

panels) were cultured with Csf-1 alone; Csf-1 plus OPGL; Csf-1 plus supernatant of activated T cells (csf-1/sup-T); Csf-1 plus fixed OPGL⁺CD4⁺ T cells (csf-1/para-T); or Csf-1 plus fixed OPGL⁺CD4⁺ T cells plus OPG (500 ng ml⁻¹) (csf-1/para-T/OPG). Arrowheads show TRAP⁺ osteoclasts. **e–g**, Calcitonin receptor (CTR), RANK, cathepsin K and β_3 -integrin expression. Haematopoietic progenitors from *opgl*^{-/-} mice were cultured as above. CTR, RANK and β -actin mRNA expression were detected by RT-PCR (**e**). Expression of cathepsin K (**f**) and β_3 -integrin (**g**) was detected by immunostaining. Arrows indicate osteoclasts.

Table 1 Bone mineral density

Groups	BMD in metaphysis (mg cm ⁻³)		BMD in diaphysis (mg cm ⁻³)	
	Total	Trabecular	Total	Trabecular
<i>rag1</i> ^{-/-}	363.3 ± 14.5	309.5 ± 18.5	644.4 ± 3.6	ND
<i>ctla4</i> ^{+/-} <i>rag1</i> ^{-/-}	379.6 ± 7.3	291.5 ± 15.6	626.4 ± 17.0	462.8 ± 19.6
<i>ctla4</i> ^{-/-} <i>rag1</i> ^{-/-}	302.7 ± 14.9*	209.5 ± 44.9*	529.8 ± 3.4*	353.0 ± 10.8*

Bone mineral density (BMD) was measured by peripheral quantitative computed tomography of tibial bones from *rag1*^{-/-} (*n* = 3) and *rag1*^{-/-} mice reconstituted with *ctla4*^{+/-} (*n* = 3) or *ctla4*^{-/-} (*n* = 3) bone marrow cells eight weeks after cell transfer.
* Statistically significant difference between different groups (Student's *t*-test: *P* < 0.05). Values are given as the mean ± s.d. ND, not determined.

density and significantly reduced the number of osteoclasts at the growth plates (Table 2, and Fig. 2k, l). These results show that systemic activation of T cells leads to bone loss *in vivo* through OPGL.

Arthritis in humans is characterized by synovial inflammation, erosion of bone and cartilage, severe joint pain and ultimately life-long crippling¹⁵. In Lewis rats, experimental induction of adjuvant-induced arthritis (AdA) leads to severe inflammation in the bone marrow and soft tissues surrounding joints accompanied by extensive local bone and cartilage destruction, loss of bone mineral density and crippling²⁰. This condition in rats mimics many of the clinical and pathological features of human rheumatoid arthritis. Lesions in AdA are strictly dependent on T-cell activation²¹. Thus, we tested the role of OPGL in the pathogenesis of AdA arthritis in Lewis rats. OPGL protein was expressed on the surface of activated T cells isolated from rats showing clinical onset of arthritis, and OPGL mRNA could be detected in both synovial cells and inflammatory cells by *in situ* hybridization. At the onset of disease, arthritic rats

Table 2 Numbers of osteoclasts in the cortical bone

Groups	Number of osteoclasts in bone area	Number of osteoclasts in total area	BA%TA
<i>ctla4</i> ^{+/-}	7.69 ± 2.30	1.82 ± 0.73	23.5 ± 3.4
<i>ctla4</i> ^{-/-}	77.69 ± 23.12*	11.63 ± 3.85*	14.4 ± 1.7*
<i>ctla4</i> ^{-/-} /OPG	33.52 ± 12.18†	4.00 ± 1.37†	ND
<i>rag1</i> ^{-/-}	6.86 ± 3.43	2.58 ± 1.33	35.9 ± 5.6
<i>ctla4</i> ^{+/-} <i>rag1</i> ^{-/-}	18.81 ± 13.12	7.74 ± 6.09	33.7 ± 5.4
<i>ctla4</i> ^{-/-} <i>rag1</i> ^{-/-}	167.06 ± 32.59‡	26.43 ± 1.80‡	16.6 ± 1.7‡

Osteoclast numbers (mean ± s.e.m.; see Methods) were counted in tibial bones from *ctla4*^{+/-} (*n* = 3), *ctla4*^{-/-} (*n* = 3) and *ctla4*^{-/-} mice injected subcutaneously daily with OPG (1 mg kg⁻¹ body weight in PBS) on days 21–29 after birth; and *rag1*^{-/-} (*n* = 3) and *rag1*^{-/-} mice reconstituted with *ctla4*^{+/-} (*n* = 3) or *ctla4*^{-/-} (*n* = 3) bone marrow cells eight weeks after cell transfer. Histomorphometry was performed on cortical bone in the diaphysis of tibial bones and is shown as mean ± s.d. of bone area (BA) as a percentage of total tissue area (TA). ND, not determined.
* Statistically significant differences between *ctla4*^{+/-} and *ctla4*^{-/-} groups.
† Statistically significant differences between *ctla4*^{-/-} and OPG-treated *ctla4*^{-/-} groups.
‡ Statistically significant differences between *ctla4*^{+/-}*rag1*^{-/-} and *ctla4*^{-/-}*rag1*^{-/-} groups (Student's *t*-test; *P* < 0.05).

were either left untreated or treated with OPG for seven days. Inhibition of OPGL through OPG had no effect on the severity of inflammation (Fig. 3a); however, OPG treatment completely abolished the loss of mineral bone density (Fig. 3b). Histologically, OPG-treated arthritic rats exhibited minimal loss of cortical and trabecular bone, whereas untreated AdA animals developed severe bone lesions characterized by partial to complete destruction of cortical and trabecular bone, erosion of the articular cartilages and clinical crippling (Fig. 4a–f). Bone destruction in untreated arthritic rats correlated with a dramatic increase in osteoclast numbers, whereas OPG treatment prevented the accumulation of osteoclasts (Figs 3c and 4e, f). These results show that OPGL is a key mediator of joint destruction and bone loss in adjuvant arthritis.

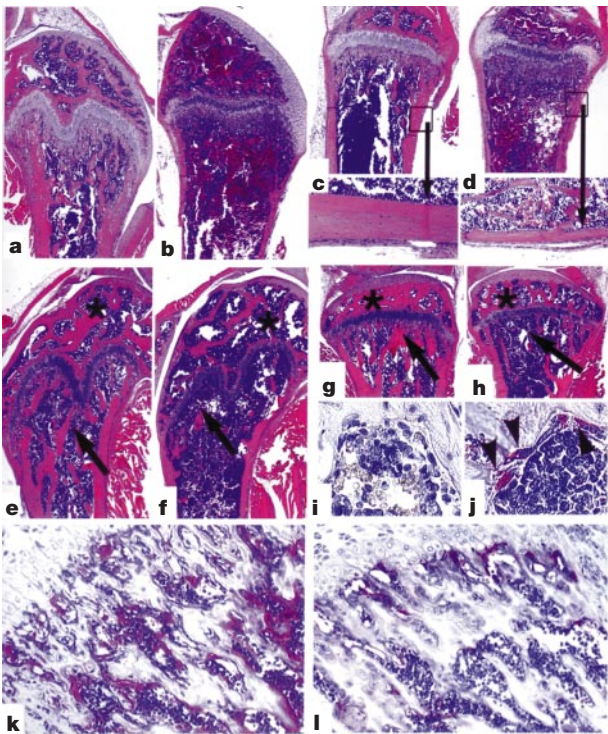


Figure 2 Activated T cells affect bone physiology *in vivo*. **a–d**, Histology of femur (**a, b**) and tibia (**c, d**) of *ctla4*^{+/-} (**a, c**) and *ctla4*^{-/-} (**b, d**) mice. Note significantly reduced thickness of cortical bone in *ctla4*^{-/-} mice. **e–h**, Histology of femur (**e, f**) and tibia (**g, h**) from *rag1*^{-/-} mice reconstituted with *ctla4*^{+/-} (**e, g**) and *ctla4*^{-/-} (**f, h**) bone marrow cells. Samples were analysed eight weeks after cell transfer. *ctla4*^{-/-}*rag1*^{-/-} mice show extended resorption of trabecular bone below the growth plate (arrows) and in

the epiphysis (asterix) of both femur and tibia. **i, j**, Increased number of TRAP⁺ osteoclasts in *ctla4*^{-/-}*rag1*^{-/-} (**j**) compared with *ctla4*^{+/-}*rag1*^{-/-} chimaeras (**i**). **k, l**, OPG treatment reduces osteoclast numbers in *ctla4*^{-/-} mice. *ctla4*^{-/-} mice were injected subcutaneously daily with PBS (**k**) or OPG (1 mg kg⁻¹ body weight in PBS) (**l**) on days 21–29 after birth. Note the decreased number of TRAP⁺ osteoclasts in the femur growth plate of OPG-treated *ctla4*^{-/-} mice (**l**) (TRAP, 20×). Staining: HE (**a–h**); TRAP (**i–l**).

Alteration of cartilage structures leading to cartilage collapse constitutes a critical step in arthritic joint destruction. There is controversy about whether cartilage destruction occurs independently of bone loss, or whether damage to the subchondral bone indirectly causes cartilage deterioration^{22,23}. Because OPG treatment prevented bone loss but not inflammation, we analysed the structural integrity of joint cartilage. In untreated arthritic rats, partial or complete erosion of the cartilage matrix in both the central and peripheral regions of joint surfaces was observed, as assessed by loss of toluidine blue staining (Fig. 4h). These erosions occurred in association with destruction of the underlying bony plate. In striking contrast, the integrity of cartilage was preserved in OPG-treated arthritic rats (Fig. 4i). Whether OPG protects the cartilage indirectly by maintain the underlying subchondral bone or has a direct role in cartilage maintenance needs to be determined. In addition, we have found that T cells isolated from joints of all

human rheumatoid arthritis ($n = 20$) and all osteoarthritis ($n = 10$) patients express OPGL (see Supplementary Information). The significance of T cell-derived OPGL in human arthritis needs to be determined. These data show that OPGL is a key regulator of bone metabolism, and that inhibition of OPGL activity by OPG can prevent cartilage destruction, a critical, irreversible step in the pathogenesis of arthritis.

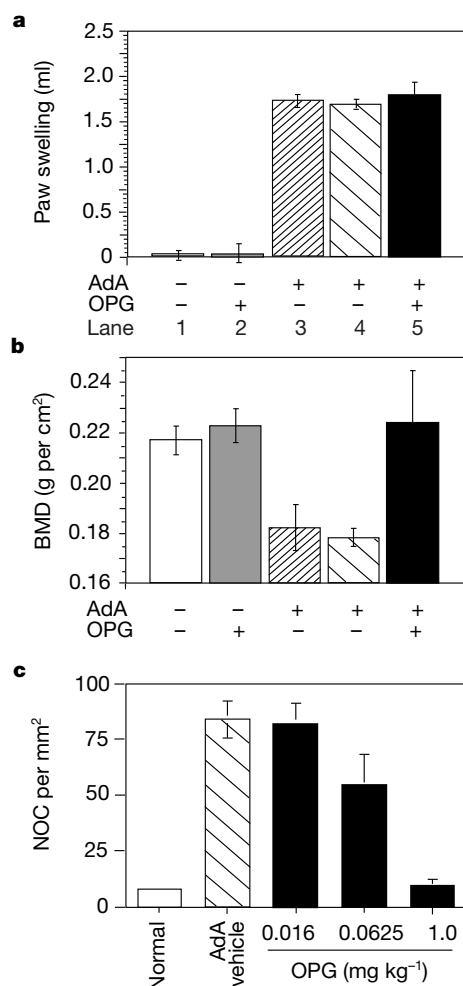


Figure 3 OPG blocks bone loss in adjuvant-induced arthritis (AdA). **a**, Severity of inflammation was calculated by measuring the volume of hind paw swelling on day 16 after initiation of disease. Lane 1, untreated controls; lane 2, OPG ($1 \text{ mg kg}^{-1} \text{ day}^{-1}$) treatment but no AdA; lane 3, AdA but no OPG treatment; lane 4, AdA plus OPG vehicle control; lane 5, AdA plus OPG ($1 \text{ mg kg}^{-1} \text{ day}^{-1}$). Mean values of swelling (water displacement (ml)) $\pm$ s.d. are shown ($n = 6$ per group). Paw swelling was similar between AdA and AdA plus OPG groups at all time points (days 9–15). **b**, Bone mineral density (BMD) of the tibiotarsal region was determined on day 16. Lanes are as in **a**. Mean values of BMD $\pm$ s.d. are shown ($n = 6$ per group). **c**, Numbers of osteoclasts (NOC) per mm^2 of total bone area in the navicular tarsal bone were determined on day 16. For induction of AdA in Lewis rats and daily subcutaneous injections with soluble OPG from day 9 (onset of disease) to day 15, see Methods.

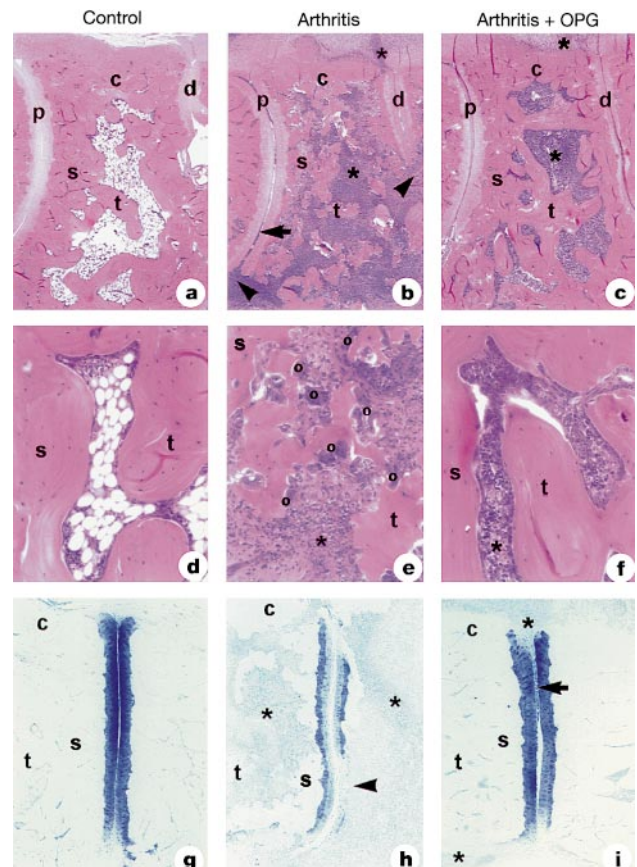


Figure 4 OPG prevents bone and cartilage destruction even in the presence of severe inflammation. **a, d**, Bone and joint structure in the normal hind paw showing dense bony plates, intact articular cartilages and marrow cavities containing scattered haematopoietic precursor cells. Proximal (p) distal (d) intertarsal joints. **b, e**, Disrupted bone and joint structure in AdA rats (day 16) with severe mononuclear infiltration in the bone marrow (asterisk), and pannus (arrow); advanced destruction (arrowheads) of cortical (c), subchondral (s) and trabecular (t) bone; and erosion of the articular cartilages. The marrow cavity contains a marked mononuclear cell infiltration (asterisk) containing numerous osteoclasts (o). These rats show severe clinical crippling. **c, f**, Preserved bone and joint structure of AdA rats (day 16) treated with OPG ($1 \text{ mg kg}^{-1} \text{ day}^{-1}$) on days 9–15. OPG-treated rats exhibit extensive mononuclear cell infiltration of bone marrow (asterisk) and pannus formation (arrow), but cortical (c), subchondral (s) and trabecular (t) bone and articular cartilages are intact. Note the absence of osteoclasts as compared with non-OPG-treated rats (**e**). These rats do not show any clinical signs of crippling. **g**, Normal cartilage integrity in control rats as determined by toluidine blue staining. The matrix of normal articular cartilage is uniformly stained, and the underlying bony plates are dense. **h**, Cartilage matrix degeneration in AdA rats (day 16). Uniform pallor (small arrow) in the upper half of articular cartilages denotes extensive loss of matrix proteoglycans. Subchondral (s), cortical (c) and trabecular (t) bone is extensively eroded, and the marrow cavity is filled with inflammatory cells (asterisks). **i**, OPG treatment preserves matrix proteoglycans of arthritic rats (day 16) treated with OPG ($1 \text{ mg kg}^{-1} \text{ day}^{-1}$) on days 9–15. Note modest reduction of toluidine blue staining in peripheral cartilage regions that are in direct contact with inflamed synovial tissue (asterisks) or pannus (arrows). Subchondral (s), cortical (c) and trabecular (t) bone is intact. Staining: HE (**a–c**); toluidine blue (**g–i**). Magnifications: $50\times$ (**a–c**); $250\times$ (**d–f**); $75\times$ (**g–i**).

Our results show that activated T cells can regulate systemic and local bone loss through OPGL. In summary, activated T cells produce OPGL and can directly trigger osteoclastogenesis *in vitro*; activated T cells from *ctla4*^{-/-} mice have a destructive effect on bone mineral density *in vivo* that can be reversed through inhibition of OPGL; and inhibition of OPGL through OPG can completely prevent bone and cartilage loss in a T-cell-dependent arthritis model. In addition, we have found that six out of seven spontaneously arising T-cell lymphomas in mice express high levels of OPGL on T-cell surfaces, and that these lymphoma cells can directly support *in vitro* osteoclastogenesis (data not shown). As mutant mice lacking T and B cells exhibit normal bone morphology and bone mineral densities, T cells are probably not required for normal bone homeostasis. Local inflammation within the bone due to metastasis, infections or fractures, or joint inflammation in arthritis, however, attracts T cells which appear to actively participate in bone remodelling through production of OPGL. To our knowledge, our results provide the first definitive linkage between activated T cells and bone homeostasis and that systemic or local activation of T cells triggers bone loss through expression of OPGL. These findings provide the molecular explanation for bone loss associated with diseases having immune-system involvement, such as adult and childhood leukaemias, autoimmunity and various viral infections such as hepatitis and HIV. Thus, inhibition of OPGL function might ameliorate many osteopenic conditions and prevent the bone destruction and cartilage damage that ultimately cause crippling in arthritis. □

Methods

T-cell activation and osteoclastogenesis

Purified (>98% CD3⁺) CD8⁺ and CD4⁺ T cells (10⁶ per well) were seeded in Iscove's modified Dulbecco medium (IMDM) and stimulated with anti-CD3ε (2 μg ml⁻¹) and anti-CD28 (200 ng ml⁻¹) antibodies (PharMingen) in the absence or presence of the MAPK/ERK inhibitor PD98059 (PD, 4 μM), the calcineurin inhibitor cyclosporin A (CsA, 100 ng ml⁻¹), the PKC inhibitor RO-31-8220 (RO, 2 μM) or the P13K inhibitor wortmannin (WT, 2 μM) for various times. OPGL mRNA expression was detected by RT-PCR as described¹⁷. Membrane-bound OPGL was detected by double staining using OPG-fluorescein isothiocyanate (FITC) and anti-CD4-PE or anti-CD8-PE. For CD25 (IL-2Rα chain) expression, cells were stained with anti-CD25-FITC and anti-CD4-PE or anti-CD8-PE. Viable cells were collected by FACS analysis. For detection of secreted OPGL protein, T cells (5 × 10⁵) were stimulated with anti-CD3ε (2 μg ml⁻¹) plus anti-CD28 (200 ng ml⁻¹) for 2 days and metabolically labelled with 1 μCi ml⁻¹ of [³⁵S]methionine for the last 100 min of incubation. Labelled OPGL was immunoprecipitated from supernatants using either 1 μg ml⁻¹ of human OPG-Fc or 1 μg of isotype-matched control IgG. In addition, secreted OPGL protein was determined in supernatants of anti-CD3ε and anti-CD28 stimulated T cells (4 days) by ELISA. OPGL expression on the surface of T cells was detectable 2 days after stimulation with anti-CD3 and anti-CD28, and maximum expression was observed at day 4 of stimulation. Throughout the whole time of the co-culture, expression of OPGL on fixed T cells was stable (data not shown). For *in vitro* osteoclastogenesis, purified CD4⁺ T cells were stimulated for 4 days with anti-CD3ε plus anti-CD28 and fixed with 2.5% paraformaldehyde. Fixed T cells (10⁶ cells per well) were co-cultured with non-adherent hematopoietic progenitors (10⁵ cells well⁻¹) from bone marrow cells of wild-type mice or spleen cells of newborn *opgl*^{-/-} mice in the presence of 30 ng ml⁻¹ colony-stimulating factor-1 (Csf-1; Genzyme) in a flat-bottom 96-well plate. The soluble decoy receptor OPG (500 ng ml⁻¹) was added to control co-cultures to confirm that the observed effects were mediated by OPGL. OPG was used as described¹. Osteoclast differentiation was evaluated on day 4 by cytochemical staining for TRAP, which specifically labels osteoclasts with a red dye, immunohistochemistry to detect Cathepsin K and β₃-integrin, and by RT-PCR for calcitonin receptor (CTR) and RANK expression. PCR primers were β-actin sense, CACTGCCGATCCTCTCTCTC; β-actin antisense, GCTGTGCGCTTCACCGTTCCA; calcitonin receptor sense, ACAAAGCTGTGGCTGAGTG; calcitonin receptor antisense, GAAGCAGTAGATAGTGCCCA; RANK sense, GCAACCTCCAGTCAGCA; RANK antisense, GAAGTCACAGCCCTCAGAATC. To detect expression of β₃-integrin and cathepsin K protein, *opgl*^{-/-} spleen cells were cultured with wild-type T cells for 4 days, and the expression was detected as described²⁴.

In vivo bone remodelling

ctla4^{-/-} and *rag1*^{-/-} gene-targeted mice were back-crossed onto a C57BL/6 background for at least eight generations^{18,19}. For chimaeras, bone marrow cells or purified lymph-node T cells from *ctla4*^{-/-} and *ctla4*^{+/+} littermate mice were transferred into 6-week-old irradiated (300 rad) *rag1*^{-/-} and into 6-week-old irradiated (500 rad) *opgl*^{-/-} mice. Chimaerism was monitored by FACS staining for labelled monoclonal antibodies reactive to surface markers on T cells (anti-CD4, anti-CD8, anti-TCR) or B cells (anti-CD19, anti-B220, anti-slgM) (all from PharMingen). In all experiments, chimaerism of these mice was more than

90%. RAG-chimaeric mice were necropsied 8–10 weeks after cell transfer. *ctla4*^{-/-} T cell → *opgl*^{-/-} chimaeras were necropsied 12 days after cell transfer. Bone mineral density was determined as described². Histology and histomorphometry were done as described². Osteoclast numbers were determined in a 0.5 × 1.0 mm rectangle area distal to the growth plate in the tibia. The field of measurement started adjacent to the growth plate, in the centre of the marrow cavity not including the cartilage or cortical bone.

Adjuvant-induced arthritis (AdA)

Eight- to nine-week male Lewis rats were given subcutaneous (s.c.) injections of *Mycobacterium tuberculosis* (0.5 mg emulsified in 50 μl of paraffin oil) into the base of the tail (Difco Laboratories, Detroit). At the onset of clinical inflammation (day 9 after inoculation), animals were injected s.c. with soluble OPG (0.016, 0.0625 and 1 mg kg⁻¹ day⁻¹) on each of days 9–15. On day 16, animals were killed and bone mineral density (BMD) of the tibiotarsal region was determined by DEXAScan of the hind paws using a dual-energy X-ray absorptiometer (Hologic QDR 4500). Severity of inflammation was monitored daily on days 8–16 by measuring the volume of hind paws by water displacement. Paws were processed into paraffin as described above, and serial sections were stained with haematoxylin and eosin (HE) (to grade inflammatory and bone destructive changes) and toluidine blue (to assess the integrity of cartilage by visualizing matrix proteoglycans). Criteria for scoring the extent of inflammation and cartilage matrix integrity have been defined²⁰. Numbers of osteoclasts were counted in the navicular tarsal bone using histomorphometric measurements by tracing the section image onto a digitizing plate with the aid of a camera lucida and 'Osteomeasure' (Osteometrics Inc., Decatur, GA) bone analysis software. Two separate 560 × 560 μm regions of the talus bone were measured for each individual section. Osteoclasts included in the measurement were identified morphologically and had to be in contact with a bone surface. Experimental results are reported as the average number of osteoclasts (NOC) per mm² tissue of both left and right ankles. Animal experiments were in accordance with institutional guidelines.

Received 6 August; accepted 23 September 1999.

- Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
- Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165–176 (1998).
- Kong, Y. Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
- Anderson, D. M. *et al.* A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
- Wong, B. R. *et al.* TRANCE (tumor necrosis factor [TNF]-related activation-induced cytokine), a new TNF family member predominantly expressed in T cells, is a dendritic cell-specific survival factor. *J. Exp. Med.* **186**, 2075–2080 (1997).
- Hsu, H. *et al.* Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation and activation induced by osteoprotegerin ligand. *Proc. Natl Acad. Sci. USA* **96**, 3540–3545 (1999).
- Wong, B. R. *et al.* TRANCE is a novel ligand of the tumor necrosis factor receptor family that activates c-Jun N-terminal kinase in T cells. *J. Biol. Chem.* **272**, 25190–25194 (1997).
- Felix, R., Hofstetter, W. & Cecchini, M. G. Recent developments in the understanding of the pathophysiology of osteoporosis. *Eur. J. Endocrinol.* **134**, 143–156 (1996).
- Roodman, G. D. Advances in bone biology: the osteoclast. *Endocr. Rev.* **17**, 308–332 (1996).
- Roodman, G. D. Paget's disease and osteoclast biology. *Bone* **19**, 209–212 (1996).
- Coleman, R. E., Smith, P. & Rubens, R. D. Clinical course and prognostic factors following bone recurrence from breast cancer. *Br. J. Cancer* **77**, 336–340 (1998).
- Stellon, A. J., Davies, A., Compston, J. & Williams, R. Bone loss in autoimmune chronic active hepatitis on maintenance corticosteroid therapy. *Gastroenterology* **89**, 1078–1083 (1985).
- Oliveri, M. B., Mautalen, C. A., Rodriguez Fuchs, C. A. & Romanelli, M. C. Vertebral compression fractures at the onset of acute lymphoblastic leukemia in a child. *Henry Ford Hosp. Med. J.* **39**, 45–48 (1991).
- Piepkorn, B. *et al.* Bone mineral density and bone metabolism in diabetes mellitus. *Horm. Metab. Res.* **29**, 584–591 (1997).
- Feldmann, M., Brennan, F. M. & Maini, R. N. Role of cytokines in rheumatoid arthritis. *Annu. Rev. Immunol.* **14**, 397–440 (1996).
- Kotake, S. *et al.* IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. *J. Clin. Invest.* **103**, 1345–1352 (1999).
- Josien, R., Wong, B. R., Li, H. L., Steinman, R. M. & Choi, Y. TRANCE, a TNF family member, is differentially expressed on T cell subsets and induces cytokine production in dendritic cells. *J. Immunol.* **162**, 2562–2568 (1999).
- Waterhouse, P. *et al.* Lymphoproliferative disorders with early lethality in mice deficient in CtlA-4. *Science* **270**, 985–988 (1995).
- Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
- Bendele, A. *et al.* Efficacy of sustained blood levels of interleukin-1 receptor antagonist in animal models of arthritis: comparison of efficacy in animal models with human clinical data. *Arthritis Rheum.* **42**, 498–506 (1999).
- Panayi, G. S., Lanchbury, J. S. & Kingsley, G. H. The importance of the T cell in initiating and maintaining the chronic synovitis of rheumatoid arthritis. *Arthritis Rheum.* **35**, 729–735 (1992).
- Muller-Ladner, U., Gay, R. E. & Gay, S. Molecular biology of cartilage and bone destruction. *Curr. Opin. Rheumatol.* **10**, 212–219 (1998).
- Conway, J. G. *et al.* Inhibition of cartilage and bone destruction in adjuvant arthritis in the rat by a matrix metalloproteinase inhibitor. *J. Exp. Med.* **182**, 449–457 (1995).
- Faust, J. *et al.* Osteoclast markers accumulate on cells developing from human peripheral blood mononuclear precursors. *J. Cell. Biochem.* **72**, 67–80 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. C. Keystone for providing patient samples and C. Dunstan for critical comments. Technical assistance was provided by Y. Cheng, E. Julian, C. Burgh, A. Shahinian and D. Duryea. We are grateful to M. E. Saunders for scientific editing and A. Hessel, A. Oliveira dos Santos, K. Bachmaier, T. Sasaki and all other members of the laboratory for comments.

Correspondence and requests for materials should be addressed to J.M.P. (e-mail: jpenning@amgen.com).

The p66^{shc} adaptor protein controls oxidative stress response and life span in mammals

Enrica Migliaccio^{*†}, Marco Giorgio^{*†}, Simonetta Mele[‡],
Giuliana Pelicci^{*}, Paolo Reboldi[§], Pier Paolo Pandolfi^{||},
Luisa Lanfranccone^{*} & Pier Giuseppe Pelicci^{*†}

^{*} Department of Experimental Oncology, European Institute of Oncology, 20141 Milan, Italy

^{||} Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, Molecular Biology and Cell Biology Programs, Sloan-Kettering Institute, New York, New York 1021, USA

[§] Istituto di Patologia Medica, Perugia University, 06100 Perugia, Italy

[‡] Istituto di Medicina Interna e Scienze Oncologiche, Perugia University, 06100 Perugia, Italy

[†] These authors contributed equally to this work

Gene mutations in invertebrates have been identified that extend life span and enhance resistance to environmental stresses such as ultraviolet light or reactive oxygen species¹. In mammals, the mechanisms that regulate stress response are poorly understood and no genes are known to increase individual life span. Here we report that targeted mutation of the mouse p66^{shc} gene induces stress resistance and prolongs life span. p66^{shc} is a splice variant of p52^{shc}/p46^{shc} (ref. 2), a cytoplasmic signal transducer involved in the transmission of mitogenic signals from activated receptors to Ras³. We show that: (1) p66^{shc} is serine phosphorylated upon treatment with hydrogen peroxide (H₂O₂) or irradiation with ultraviolet light; (2) ablation of p66^{shc} enhances cellular resistance to apoptosis induced by H₂O₂ or ultraviolet light; (3) a serine-phosphorylation defective mutant of p66^{shc} cannot restore the normal stress response in p66^{shc} cells; (4) the p53 and p21 stress response is impaired in p66^{shc} cells; (5) p66^{shc} mice have increased resistance to paraquat and a 30% increase in life span. We propose that p66^{shc} is part of a signal transduction pathway that regulates stress apoptotic responses and life span in mammals.

The mammalian proto-oncogene SHC locus encodes three proteins with relative molecular masses (*M_r*) of 52K(p52^{shc}), 46K(p46^{shc}) and 66K(p66^{shc}), which share a Src-homology2 (SH2) domain, a collagen-homology (CH1) region and a phosphotyrosine-binding (PTB) domain. p66^{shc} contains a unique amino-terminal region (CH2; Fig. 1a)^{2,3}. Like p52^{shc}/p46^{shc}, p66^{shc} becomes tyrosine phosphorylated upon activation of growth factor receptors and forms stable complexes with Grb2, an adaptor protein for the Ras exchange factor SOS²⁻⁵. However, it does not affect mitogen-activated protein kinase (MAPK) activity and inhibits *c-fos* promoter activation², indicating that p66^{shc} may not be involved in Ras activation. *c-fos* is transcriptionally activated in response to environmental stresses such as DNA-damaging agents (for example, ultraviolet light (UV)) or agents that induce oxidative damage by increasing intracellular levels of reactive oxygen species (ROS) (such as H₂O₂)⁶.

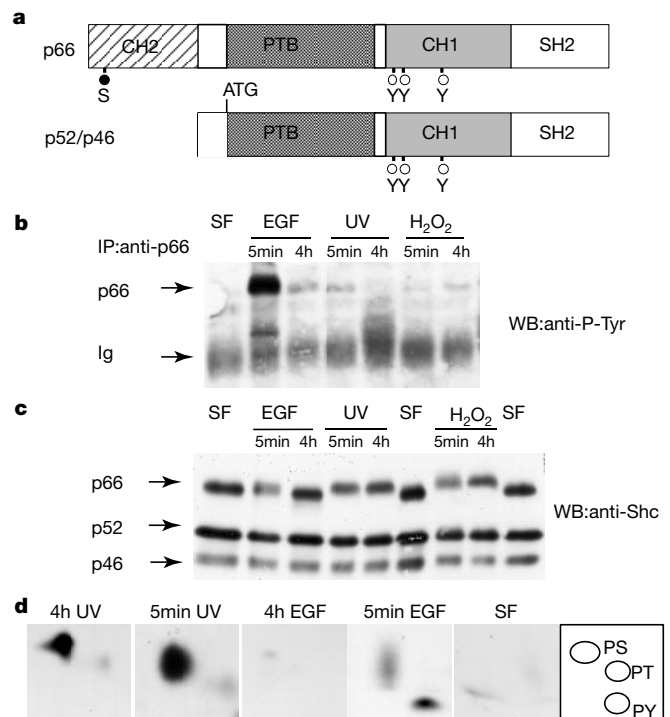


Figure 1 Serine phosphorylation of p66^{shc} by oxidative damage. **a**, Organization of Shc proteins. p66^{shc} and p52^{shc}/p46^{shc} are encoded by two distinct transcripts that differ in the use of 5' coding exons². S: S36, the major serine-phosphorylation site of p66^{shc}; Y: tyrosine-phosphorylation sites. **b**, Tyrosine phosphorylation of p66^{shc}. Anti-p66 immunoprecipitates (IP) from starved or EGF, H₂O₂ or UV-treated MEFs were blotted (WB) with anti-phosphotyrosine (anti-P-Tyr) antibodies. Ig: immunoglobulin cross-reactive polypeptides. **c**, Western blot analysis of Shc expression. Lysates from serum starved (SF) or treated MEFs were probed with anti-Shc antibodies. **d**, Phospho-amino-acid analysis of p66^{shc}. Serum-starved MEFs were labelled with [³²P]orthophosphate (SF) and treated with EGF or UV for 5 min or 4 h. Phospho-amino-acid analysis was performed on the immunoprecipitated p66^{shc} polypeptides. Positions of phosphoserine (PS), phosphothreonine (PT) and phosphotyrosine (PY) markers are indicated.

onmental stresses such as DNA-damaging agents (for example, ultraviolet light (UV)) or agents that induce oxidative damage by increasing intracellular levels of reactive oxygen species (ROS) (such as H₂O₂)⁶.

To investigate the role of p66^{shc} in stress responses, we analysed the extent of its tyrosine phosphorylation in mouse embryo fibroblasts (MEFs) treated with UV, H₂O₂ or epidermal growth factor (EGF). EGF, but not UV or H₂O₂, induced tyrosine phosphorylation of p66^{shc}, which was maximal after 5 min (Fig. 1b). However, UV and H₂O₂ provoked marked gel retardation of p66^{shc} (Fig. 1c), indicating other possible post-translational modifications. Phospho-amino-acid analysis of p66^{shc} phosphorylation after treatment with UV (Fig. 1d) or H₂O₂ (data not shown) revealed a marked and persistent (up to 4 h) increase in phosphoserine. It appears, therefore, that p66^{shc} is involved in the cellular responses to both growth factors and environmental stresses; possibly through distinct pathways, as EGF induced phosphorylation on tyrosine, whereas UV or H₂O₂ induced phosphorylation on serine.

To investigate the function of p66^{shc} in the environmental stress response, we analysed the effects of ablation of p66^{shc} on the cellular response to oxidative (H₂O₂) or DNA (UV or γ -irradiation) damage. MEFs were derived from mice carrying a targeted mutation of the *Shc* CH2 exon, which did not affect p52^{shc}/p46^{shc} coding sequences (M.G. *et al.*, in preparation). As expected, p66^{shc} was reduced in p66^{shc} and absent in p66^{shc} MEFs, whereas p52^{shc}/

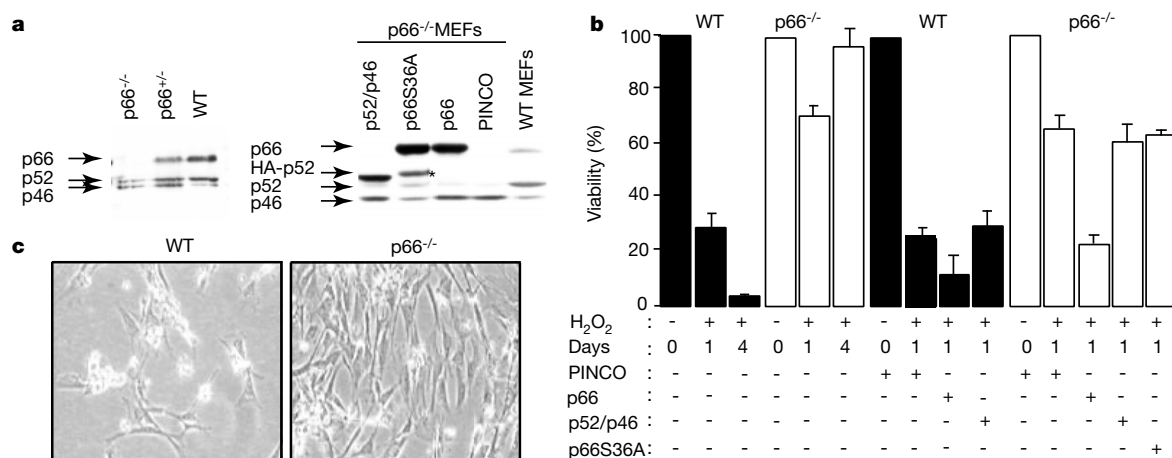


Figure 2 Targeted mutation of p66^{shc} induces cellular resistance to oxidative damage. **a**, Anti-Shc western blots of wild-type (WT), p66^{shc/+} or ^{-/-} MEFs (left) and of p66^{shc/-} MEFs infected with PINCO or PINCO retroviruses expressing p66^{shc}, p52^{shc}/p46^{shc} or p66^{shc}S36A cDNAs (right). **b**, Viability of MEFs after H₂O₂ treatment. Wild-type and p66^{shc/-} MEFs were treated with H₂O₂ for 1 or 4 days, as indicated. In separate experiments, cells were infected with PINCO or with PINCO retroviruses expressing the

p66^{shc}, p52^{shc}/p46^{shc} or p66S36A cDNAs, kept for 48 h to allow gene expression of exogenous cDNAs and treated with H₂O₂. Cell viability was determined by trypan blue exclusion. Results are expressed as percentage of viable cells with respect to untreated controls and represent the mean of triplicate cultures. This experiment is representative of four that gave similar results. **c**, Morphology of wild-type and p66^{shc/-} MEFs 24 h after H₂O₂ treatment (original magnification 100×).

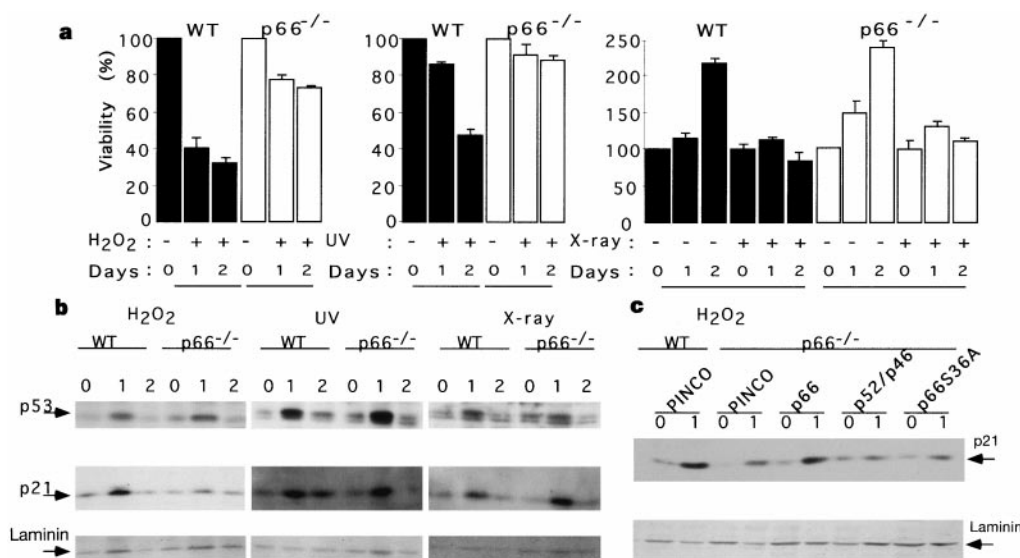


Figure 3 p53 and p21 regulation in p66^{shc/-} MEFs after H₂O₂, UV and X-rays. **a**, MEFs were treated with H₂O₂, UV or X-rays and evaluated by trypan blue exclusion. **b**, Western blot analysis of p53, p21 and laminin expression in wild-type and p66^{shc/-} MEFs after H₂O₂, UV and X-rays. The cellular lysates were derived from the experiment shown in **a**.

c, Western blot analysis of p21 and laminin expression in PINCO-infected wild-type MEFs and in p66^{shc/-} cells after infection with PINCO retroviruses expressing p66^{shc}, p52^{shc}/p46^{shc} or p66S36A.

p66^{shc} was detectable in both (Fig. 2a, left). Wild-type (WT) MEFs were sensitive to H₂O₂, with 70% of cells dying after one day of treatment (Fig. 2b, c). *In situ* terminal deoxynucleotidyl transferase assay revealed massive apoptosis (>70%; data not shown). Overexpression of p66^{shc} increased the susceptibility of wild-type MEFs (85–90% cell death; Fig. 2b), whereas p66^{shc} ablation increased resistance of MEFs to H₂O₂ (<30% cell death; Fig. 2b, c). p66^{shc/-} MEFs that survived oxidative stress maintained their proliferative potential, as revealed by their ability to restore the initial cell number within a few days (Fig. 2b). Similar results were obtained by treating cells with UV: in wild-type MEFs, about 80% and 40% of the cells survived after one and two days of treatment, respectively, whereas in the p66^{shc/-} MEFs the cytotoxic effect of UV was

negligible (Fig. 3a). Apoptosis after one day of treatment was 20–25% and 5–7.5% in wild-type and p66^{shc/-} MEFs, respectively (data not shown). Expression of p66^{shc}, but not p52^{shc}/p46^{shc}, into p66^{shc/-} cells restored a normal response to H₂O₂ (Fig. 2a, b) and UV (data not shown). γ -Irradiation, in contrast, induced arrest of cell proliferation (G1 arrest; data not shown) without significant loss of cell viability (Fig. 3a) in wild-type and p66^{shc/-} MEFs. UV and X-rays induce DNA damage; only UV, however, increases intracellular ROS and induces oxidative damage, which is the main causative agent in the physiological response to UV⁷. It appears, therefore, that p66^{shc} is a crucial component of the apoptotic response to oxidative damage, whereas it has no effect on the arrest of cell proliferation induced by DNA damage.

Activation of serine-threonine kinases is a prominent feature of the response to oxidative damage (H_2O_2 or UV)⁸. To test whether p66^{shc} is a cytoplasmic transducer of oxidative stress signals, we evaluated the biological activity of a serine-phosphorylation mutant of p66^{shc}. The isolated p66^{shc} CH2 region, expressed in fibroblasts, exhibited gel retardation upon treatment with EGF or H_2O_2 , suggesting modifications of phosphoserine-type (Fig. 4a, top). Alanine substitution of the CH2 serine 36 (S36A) abrogated the H_2O_2 -induced gel-mobility shift of both the isolated CH2 polypeptide (Fig. 4a, middle) and of the full-length p66^{shc} (Fig. 3a, bottom), and markedly reduced serine phosphorylation of p66^{shc} upon H_2O_2 treatment (Fig. 4b). Expression of the p66S36A mutant into p66^{shc} MEFs (Fig. 2a, right) did not restore a normal response to H_2O_2 (Fig. 2b). It appears, therefore, that serine phosphorylation of p66^{shc} is critical for the activation of apoptosis in cells exposed to oxidative damage. p66^{shc} might monitor the intracellular concentration of ROS and eliminate cells injured by oxidative damage.

The cellular stress response involves upregulation of the tumour suppressor p53 and the cyclin-dependent kinase inhibitor p21, a transcriptional target of p53 (ref. 9). The p21 response to UV or H_2O_2 is regulated through both p53-dependent and -independent pathways, whereas p21 upregulation by X-rays depends only on p53 activation^{10–12}. We investigated the regulation of p53 and p21 by H_2O_2 , UV and X-rays in wild-type and p66^{shc} MEFs. Upregulation of p53 was comparable in wild-type and p66^{shc} cells (Fig. 3b). However, upregulation of p21 was significantly reduced in the p66^{shc} cells following treatment with H_2O_2 and UV, but not X-rays (Fig. 3b), indicating that p66^{shc} interferes with p53-independent pathway(s) that control p21 upregulation following oxidative damage. Overexpression of p66^{shc}, but not of p66S36A or p52^{shc}/p46^{shc}, restored the normal p21 response to oxidative stress in p66^{shc} cells (Fig. 3c). To test the role of p21 in the increased stress resistance of p66^{shc} cells, we overexpressed p21 in wild-type and p66^{shc} cells. Overexpression of p21 protected wild-type MEFs from H_2O_2 -induced cell death and did not abrogate the survival

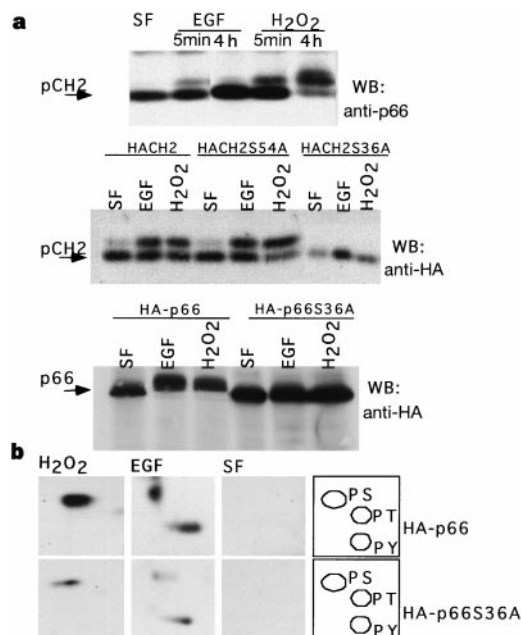


Figure 4 Mapping of p66^{shc} serine-phosphorylation sites. **a**, Wild-type MEFs were transfected with expression vectors for haemagglutinin (HA)-tagged versions of wild-type (top), S54A or S36A (middle) CH2 polypeptides, p66^{shc} or p66S36A (bottom). Serum-starved (SF) cells were treated with EGF or H_2O_2 and analysed by western blotting using anti-p66 or anti-HA antibodies, as indicated. **b**, Phospho-amino-acid analysis of HA-p66^{shc} and HA-p66S36A.

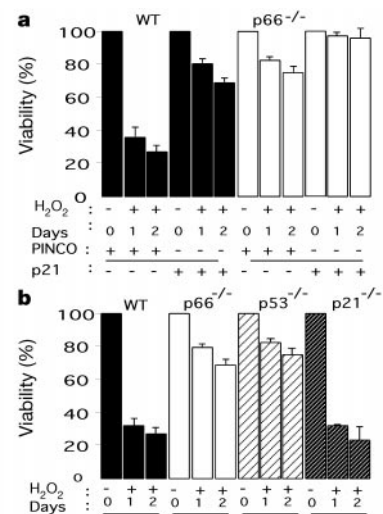


Figure 5 Effects of p21 expression on the oxidative stress response of wild-type and p66^{shc} cells. **a**, Wild-type and p66^{shc} MEFs were infected with PINCO or with PINCO retrovirus expressing p21 and treated with H_2O_2 . Cell viability was determined after 1 or 2 days by trypan blue exclusion. This experiment is representative of two that gave similar results. **b**, Viability of wild-type, p66^{shc} -/-, p53^{shc} -/- and p21^{shc} -/- MEFs after exposure to H_2O_2 .

effect of p66^{shc} ablation (Fig. 5a). p21^{shc} MEFs had a normal oxidative stress response, whereas p53^{shc} MEFs were resistant (Fig. 5b)^{13,14}. It appears, therefore, that the apoptotic response to oxidative stress is mediated by p53 and antagonized by p21. Together, these findings indicate that oxidative stress may activate two pathways: one involving p53 and leading to apoptosis, and the other involving p21, through p53-dependent and -independent mechanisms, and protecting from apoptosis. Ablation of p66^{shc} interrupts both pathways of p21 upregulation and p53-dependent apoptosis induced by oxidative stress.

To examine the ability of p66^{shc} mice to resist oxidative stress *in vivo*, animals were treated with paraquat, which generates superoxide anions upon cellular intake¹⁵. Ten mice of ten weeks of age were injected intraperitoneally with 70 mg kg⁻¹ paraquat. All of the five wild-type mice died within 48 h, whereas, of the five injected p66^{shc} mice, two died within the first 48 h, two died within 72 h and one survived for several weeks. The same experiment was repeated using five wild-type and five p66^{shc} mice of nine weeks of age, with comparable results. Cumulative statistical analysis revealed a 40% increase in the mean survival time of the p66^{shc} mice over the wild type (67.4 ± 6.7 and 44.8 ± 4.6 h, respectively; $P = 0.0109$) (Fig. 6a). These results indicate that ablation of p66^{shc} expression enhances resistance to oxidative stress *in vivo*.

Enhanced resistance to environmental stress correlates with increased life span in invertebrates. In *Saccharomyces cerevisiae*, deletions of *RAS1* (ref. 16) or mutations in the *SIR4* (ref. 17) locus increase life span and stress resistance. In *Caenorhabditis elegans*, mutants of *clk-1* and of the genes of the dauer/insulin receptor signalling pathway (*daf-2*; *age-1*, *AKT1* and *AKT-2*; *PDK1*) survive longer and are more resistant to ROS, heat and UV^{18,19}. In *Drosophila melanogaster*, selection for late-life fitness is associated with greater resistance to environmental stresses²⁰, and hypomorphic mutants of the *mth* locus live 35% longer and are more resistant to dietary paraquat and starvation²¹. A few progeric mutations have been described in *C. elegans* (Mev-1) and *Drosophila* (catalase or SOD), and, in all cases, decreased life span correlates with enhanced sensitivity to oxidative stress^{22,23}. We conducted a prospective observational study to evaluate the effects

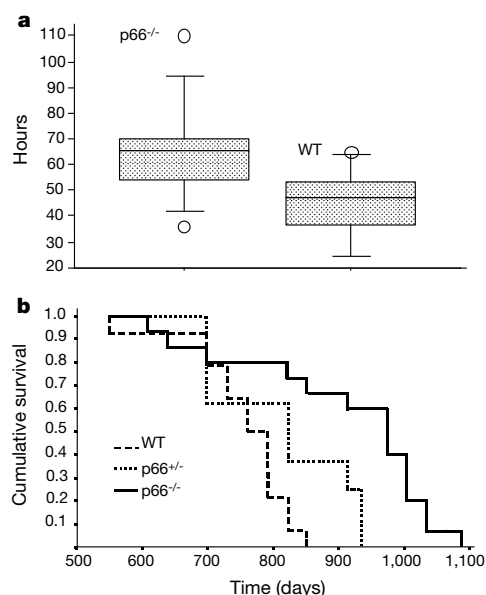


Figure 6 Increased resistance to oxidative stress and prolonged life span of p66^{shc-/-} mice. **a**, 10 wild-type and 10 p66^{shc-/-} mice were treated with paraquat in two separate experiments. In the first, paraquat was administered to 5 wild-type and 5 p66^{shc-/-} mice of 10 weeks of age; in the second, to 5 wild-type and 5 p66^{shc-/-} mice of 9 weeks of age. Cumulative statistical analysis (two-way analysis of variance) revealed representation of the mean survival time (h) of the wild-type and p66^{shc-/-} mice in a box and whisker plot is shown. **b**, Cumulative survival (Kaplan and Meier) of wild-type, p66^{shc+/-} and p66^{shc-/-} mice.

of the p66^{shc} mutation on mouse survival. Thirty-seven mice born in August 1996 from p66^{shc+/-} heterozygous parents were not sacrificed and were maintained under identical conditions. There were 14 wild-type, 8 p66^{shc+/-} and 15 p66^{shc-/-} mice. After 28 months, when all the wild-type animals had died, three of the eight heterozygous (37%) and 11 of the 15 homozygous mice (73%) were still alive. The remaining three p66^{shc+/-} and 11 p66^{shc-/-} mice died after a further two and eight months, respectively. Comparison of survival curves by the Kaplan and Meier method²⁴ showed a highly significant difference among the three groups (log-rank 16.72; $P = 0.0002$), with median survivals of 761 ± 19.02 days (wild type), 815.38 ± 37.46 (p66^{shc+/-}) and 973 ± 37.31 (p66^{shc-/-}) (Fig. 6b). Comparison of survival curves of wild-type versus p66^{shc-/-} mice showed an ~30% increase in life span of the p66^{shc-/-} mice (log-rank 13.29; $P = 0.0003$). Differences in the cumulative survival of the wild-type versus p66^{shc+/-} group, however, did not reach full statistical significance (log-rank 3.62; $P = 0.057$). The only other known model of increased life span in mammals is food restriction, which, in rodents, increases mean and maximum survival times^{25,26}. However, no statistically significant differences were found in body weight and food consumption of the p66^{shc-/-} mice. Phenotypical and histopathological analysis revealed no obvious abnormalities in the p66^{shc-/-} mice (see Supplementary Information). In conclusion, it appears that homozygous mutation of p66^{shc} increases life span in mice, in the absence of any obvious defective phenotype.

Among currently accepted evolutionary theories, it is postulated that ageing is a post-reproductive process which has escaped the force of natural selection and which evolved through selection of alleles with early life benefits combined with pleiotropically harmful effects later in life²⁷. The postulated genes, as they would be actively selected, are expected to regulate fundamental cellular processes common to different species. Accumulation of oxidative cellular damage, inflicted by ROS, is the major proximal cause of ageing in

both invertebrates and mammals. Cellular defences against oxidative damage exist (catalase and Cu/Zn superoxide dismutase, SOD); however, they are not completely efficient, as oxidatively damaged macromolecules accumulate in virtually all ageing organisms examined²⁸. In *C. elegans* and *Drosophila*, there is evidence that genes controlling ROS metabolism determine life span and that decreased ROS production might be responsible for elevated resistance to oxidative stress and increased longevity^{23,29}. In mammals, restriction of caloric intake lowers steady-state levels of oxidative stress and damage and extends the maximum life span³⁰. p66^{shc} is part of a signal transduction pathway that is activated by increased intracellular ROS and which leads to apoptosis. Mutation of this pathway increases cellular and organism oxidative stress resistance and life span. Biochemical and genetic investigation of the p66^{shc} signalling pathway should lead to better understanding of the control mechanisms and relationships between ROS metabolism, cellular survival and organism ageing in mammals. □

Methods

Cells, plasmids and antibodies

MEFs derived from wild-type 129, p66^{shc+/-} and ^{-/-}, p53^{-/-} (from A. R. Clarke), and p21^{-/-} (from M. Serrano) mice were treated with 30 ng ml⁻¹ EGF or 400 μ M H₂O₂, or irradiated with 10 J m⁻² UV-C (254 nm) or 5 Gy X-rays. The p21, p52^{shc}-p46^{shc}, p66^{shc}, p66^{shc}S36A, haemagglutinin (HA)-CH2, HA-CH2S35A, HA-CH2S54A, HA-p66^{shc}, HA-p66^{shc}S36A and HA-p66^{shc}S54A complementary DNAs were cloned into the pCDNA3 or PINCO expression vectors. The anti-Shc and anti-p66^{shc} polyclonal antibodies recognize the SH2 domain of all three Shc isoforms and the CH2 region, respectively²³. The anti-HA, anti-lamin B (Ab1), anti-p53 (CM5) and anti-p21 (F5) antibodies were purchased from Oncogene Research Products, Novo Castra and Santa Cruz, respectively.

Metabolic labelling and phospho-amino-acid analysis

Serum-starved cells were labelled for 4 h in phosphate-free DMEM containing 0.5% FBS and 1 mCi ml⁻¹ [³²P]orthophosphate, treated with EGF, H₂O₂ or UV and lysed, p66^{shc} proteins were immunoprecipitated, resolved by SDS-PAGE, transferred to PVDF membranes and hydrolysed in 6 M HCl for 60 min at 110 °C. The hydrolysis products were separated in the presence of phosphoserine, phosphothreonine and phosphotyrosine markers by SDS-PAGE at pH = 1.9 and pH = 3.5 in two dimensions on TLC plates.

Transfections and infections

MEFs were transfected with the Lipofectamine PLUS Reagent GibcoBRL (averaged transfection efficiency 50%). For retrieval infections, we used the PINCO vector, which expresses the green fluorescence protein (GFP) cDNA from an internal cytomegalovirus promoter. The efficiency of infection (~80%) was determined by FACS analysis of GFP-positive cells 48 h after infection.

Paraquat experiments

The wild-type and p66^{shc-/-} mice are carriers of mouse hepatitis virus and free of other common viral agents (Sendai, Pneumonia and Mycoplasma). Animals were maintained under a 12 h light–12 h dark cycle and given food and water *ad libitum*. For paraquat experiments, 10 wild-type and 10 p66^{shc-/-} male mice were injected intraperitoneally with 70 mg kg⁻¹ paraquat in PBS.

Received 21 July; accepted 8 September 1999.

- Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of aging: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
- Migliaccio, E. *et al.* Opposite effects of the p52^{shc}/p46^{shc} splicing isoforms on the EGF receptor-MAP kinase-fos signaling pathway. *EMBO J.* **16**, 706–716 (1997).
- Pellicci, G. *et al.* A novel transforming protein (Shc) with a SH2 domain is implicated in mitogenic signal transduction. *Cell* **70**, 93–104 (1992).
- Bonfanti, L., Migliaccio, E., Pellicci, G., Lanfrancone, L. & Pellicci, P. G. Not all Shc's roads lead to Ras. *Trends Biochem. Sci.* **21**, 257–261 (1996).
- Rozakis-Adcock, M. *et al.* Association of the Shc and Grb2/Sem5 SH2-containing proteins is implicated in activation of the Ras pathway by tyrosine kinase. *Nature* **360**, 689–692 (1992).
- Sen, C. & Packer, L. Antioxidant and redox regulation of gene transcription. *FASEB J.* **10**, 709–720 (1996).
- Renzing, J., Hansen, S. & Lane, D. P. Oxidative stress is involved in the UV activation of p53. *J. Cell Sci.* **109**, 1105–1112 (1996).
- Stevenson, M. A., Pollock, S. S., Coleman, C. N. & Calderwood, S. K. X-irradiation, phorbol esters, and H₂O₂ stimulate mitogen-activated protein kinase activity in NIH-3T3 cells through the formation of reactive oxygen intermediates. *Cancer Res.* **54**, 12–15 (1994).
- Lu, X. & Lane, D. P. Differential induction of transcriptionally active p53 following UV or ionizing radiation: defects in chromosome instability syndromes? *Cell* **75**, 765–778 (1993).
- Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
- Russo, T. *et al.* A p53-independent pathway for activation of WAF1/CIP1 expression following oxidative stress. *J. Biol. Chem.* **270**, 29386–29391 (1995).

12. Macleod, K. F. *et al.* p53-dependent and independent expression of p21 during cell growth, differentiation, and DNA damage. *Genes Dev.* **9**, 935–944 (1995).
13. Yin, Y., Solomon, G., Deng, C. & Barrett, J. C. Differential regulation of p21 by p53 and Rb in cellular response to oxidative stress. *Mol. Carcin.* **24**, 15–24 (1999).
14. Yin, Y. *et al.* Involvement of p85 in p53-dependent apoptotic response to oxidative stress. *Nature* **391**, 707–710 (1998).
15. De Haan, J., B. *et al.* Mice with a homozygous null mutation for the most abundant glutathione peroxidase, Gpx1, show increased susceptibility to the oxidative stress-inducing agents paraquat and hydrogen peroxide. *J. Biol. Chem.* **273**, 22528–22536 (1998).
16. Sun, J., Childress, A. M., Pinswasdl, C. & Jazwinski, S. Divergent roles of RAS1 and RAS2 in yeast longevity. *J. Biol. Chem.* **269**, 18638–18645 (1994).
17. Kennedy, B. K., Austriaco, N. R., Zhang, J. & Guarente, L. Mutation in the silencing gene SIR4 can delay aging in *S. cerevisiae*. *Cell* **80**, 485–496 (1995).
18. Murakami, S. & Johnson, T. E. A genetic pathway conferring life extension and resistance to UV stress in *Caenorhabditis elegans*. *Genetics* **143**, 1207–1218 (1996).
19. Larsen, P. L., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in *C. elegans*. *Genetics* **139**, 1576–1583 (1995).
20. Service, P. M., Hutchinson, E. W., MacKinley, M. D. & Rose, M. R. Resistance to environmental stress in *Drosophila melanogaster* selected for postponed senescence. *Physiol. Zool.* **58**, 380–389 (1985).
21. Lin, Y. J., Seroude, L. & Benzer, S. Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. *Science* **282**, 943–946 (1998).
22. Ishi, N. *et al.* A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* **394**, 694–697 (1998).
23. Orr, W. C. & Sohal, R. S. Extension of life-span by overexpression of superoxide dismutase and catalase in *D. melanogaster*. *Science* **263**, 1128–1130 (1994).
24. Marubini, E. & Valsecchi, M. G. *Analysing Survival Data from Clinical Trials and Observational Studies* (Wiley, New York, 1995).
25. Weindruch, R., Walford, R. L., Fligiel, S. & Guthrie, D. The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *J. Nutr.* **116**, 641–654 (1986).
26. Weindruch, R. & Walford, R. L. Dietary restriction in mice beginning at 1 year of age: effect on life-span and spontaneous cancer incidence. *Science* **215**, 1415–1417 (1982).
27. Medawar, P. B. Old age and natural death. *Modern Quarterly* **1**, 30–56 (1946).
28. Lithgow, G. J. & Kirkwood, T. B. L. Mechanisms and evolution of aging. *Science* **273**, 80 (1996).
29. Taub, J. *et al.* A cytosolic catalase is needed to extend adult life span in *C. elegans daf-1* mutants. *Nature* **399**, 162–166 (1999).
30. Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank V. Soares, P. P. DiFiore, K. Helin, L. Bonfini, I. Nicoletti and G. Della Porta for discussions; C. Matteucci, C. Casciari, M. Scanarini and G. Pelliccia for technical help; A. Ventura and A. Cicalese for contributions; and A. Ariesi for secretarial work. We also acknowledge continuous technical and intellectual support from L. Pozzi in conducting animal experiments. E.M. is the recipient of a fellowship from FIRCA. This work was supported by A.I.R.C.

Correspondence and requests for materials should be addressed to P.G.P. (e-mail: pgpellicci@ieo.it).

Structural insights into phosphoinositide 3-kinase catalysis and signalling

Edward H. Walker*, Olga Perisic*, Christian Ried*, Len Stephens† & Roger L. Williams*

* MRC Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

† The Babraham Institute, Babraham, Cambridge CB2 4AT, UK

Phosphoinositide 3-kinases (PI3Ks) are ubiquitous lipid kinases that function both as signal transducers downstream of cell-surface receptors and in constitutive intracellular membrane and protein trafficking pathways. All PI3Ks are dual-specificity enzymes with a lipid kinase activity which phosphorylates phosphoinositides at the 3-hydroxyl, and a protein kinase activity. The products of PI3K-catalysed reactions, phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P₃), PtdIns(3,4)P₂ and PtdIns(3)P, are

second messengers in a variety of signal transduction pathways, including those essential to cell proliferation, adhesion, survival, cytoskeletal rearrangement and vesicle trafficking^{1,2}. Here we report the 2.2 Å X-ray crystallographic structure of the catalytic subunit of PI3Kγ, the class I enzyme that is activated by heterotrimeric G-protein βγ subunits and Ras. PI3Kγ has a modular organization centred around a helical-domain spine, with C2 and catalytic domains positioned to interact with phospholipid membranes, and a Ras-binding domain placed against the catalytic domain where it could drive allosteric activation of the enzyme.

The mammalian PI3Ks can be divided into three classes on the basis of their structure and substrate specificity². The class I PI3Ks are receptor-regulated heterodimeric enzymes that preferentially phosphorylate PtdIns(4,5)P₂ *in vivo*. The class IA PI3Ks (consisting of p110α, p110β or p110δ catalytic subunits) associate with a p85 adaptor protein that is essential for interaction of these PI3Ks with receptor tyrosine kinases. The class IB PI3K (PI3Kγ) is activated by heterotrimeric G-protein subunits and associates with a p101 adaptor that is required for full responsiveness to Gβγ heterodimers^{3,4}. Class I PI3Ks are also activated by Ras. Class II PI3Ks are distinguished by a carboxy-terminal C2 domain and preferentially use PtdIns and PtdIns(4)P as substrates. Class III enzymes phosphorylate only PtdIns and lack the Ras-binding domain.

We have determined the structure of the catalytic subunit (residues 144–1,102) of porcine PI3Kγ. This construct contains all of the homology regions (HR) found in class I PI3Ks (HR1, HR2, HR3 and HR4) and has a catalytic activity similar to that of the full-length enzyme. The amino-terminal region missing from our construct of PI3Kγ is important for interaction with the p101 adaptor⁵, and the analogous region of PI3Kα interacts with the p85 adaptor. The enzyme has a modular structure consisting of four domains: a Ras-binding domain (RBD), a C2 domain, a helical domain and a catalytic domain (Fig. 1). The RBD, C2 and catalytic domains have folds similar to these modules in other proteins involved in signal transduction. The helical domain has a fold akin to HEAT repeat containing structures involved in protein–protein interactions.

The catalytic domain of the enzyme consists of a smaller N-terminal lobe (residues 726–883) and a larger C-terminal lobe (884–1092). The N-terminal lobe from kβ3 to kα3 and the first part of the C-terminal lobe (up to the end of kβ10) have a fold similar to protein kinases (reviewed in ref. 6), and this similarity extends to many of the details of the ATP-binding site (Fig. 2). This region is among the most conserved regions of the PI3Ks (Fig. 3). The structural similarity of PI3K to protein kinases is consistent with PI3Ks having protein kinase activity in addition to their lipid kinase activities^{7,8}. The sequence alignment in Fig. 3 shows the regions of PI3K that structurally superimpose with tyrosine protein kinase c-Src. The N-terminal lobe comprises a five-stranded antiparallel β-sheet flanked on one side by a helical hairpin (kα1–kα2) and a small two-stranded β-sheet (β1–β2), and on the other side by the kα3 helix and the C-terminal lobe. Strands kβ3–kβ7 correspond to the five-stranded β-sheet found in the protein kinases. The kβ3–kβ4 loop corresponds to the protein kinase β1–β2 loop (also known as the glycine-rich or P-loop). This loop interacts closely with the phosphates of the bound ATP, but unlike the protein kinases, it contains no glycine. Instead, the side chain of Ser 806, a residue that is conserved in all PI3Ks, interacts with the β-phosphate (Fig. 2). Lys 833 at the end of kβ5, corresponding to Lys 72 of c-AMP-dependent protein kinase, interacts with the α-phosphate of ATP. This residue is conserved in all PI3Ks and is covalently modified by wortmannin⁹. There are two metal-binding sites (Me; Fig. 2). Me I interacts with the conserved Asn 951, whereas Me II interacts with Asp 836 and Asp 964.

N- and C-terminal lobes are linked through a loop between strands kβ7 and kβ8. This loop forms the deepest wall of the ATP-

binding pocket and provides two hydrophobic contacts with the adenine moiety of the ATP. The C-terminal lobe forms part of the ATP-binding site, as well as the binding site for phospholipid substrates. The region between $\kappa\alpha 6$ and $\kappa\beta 9$ (residues 943–951) corresponds to the catalytic loop of the protein kinases; mutations of residues in this loop corresponding to Asp 946, Arg 947, Asp 950 and Asn 951 of PI3K γ abolish kinase activity of PI3Ks^{7,8}.

The C-terminal lobe contains a segment (964–988) analogous to the activation loop in the protein kinases; this loop is essential for the substrate specificity of the PI3Ks¹⁰. In the ATP–Lu³⁺ complex, much of this loop (968–982) is disordered. In the structure of an enzyme/chloramine T complex, all but two residues (Phe 975 and Leu 976) are visible, although high *B*-factors suggest that this loop is flexible. The activation loop is on the surface of the enzyme between the C-terminal helix $\kappa\alpha 12$ on one side and $\kappa\alpha 10$ on the other. We attempted to soak phospholipid analogues into PI3K γ crystals, but no substrate was evident in the electron density. Consequently, we modelled phospholipid headgroup binding, but because conformational changes probably occur in the activation loop and possibly in the C-terminal helix upon substrate binding, our model is only approximate. In the model, the headgroup is positioned in a cavity lined by the C-terminal helix $\kappa\alpha 12$, the activation loop and the catalytic loop (Fig. 4). This places the 5-phosphate of a PtdIns(4,5)P₂ adjacent to Lys 973 and the I-phosphate near Lys 807 and Lys 808. Lys 973 acting as a ligand of the 5-phosphate may explain why this residue is not found in the class II PI3Ks that do not phosphorylate phosphoinositides with a 5-phosphate. The basic residues nearest the 4-phosphate are Arg 947 and Lys 973. The specificity of the class III PI3Ks for phosphatidylinositol may be explained by their shorter activation loop, which may not leave sufficient space to accommodate a 4-phosphate at the bottom of the headgroup-binding pocket. PI3K δ autophosphorylates in a region just beyond the C-terminal helix $\kappa\alpha 12$ ¹¹, which results in enzyme inhibition probably by sterically preventing substrate binding. The proximity of the C-terminal segment to the substrate-binding site is consistent with autophosphorylation of this region.

The mechanism originally proposed for the enzymatic activity of protein kinases involved a residue acting as a general base to deprotonate the hydroxyl of the substrate and generate a nucleophile that would attack the γ -phosphate of ATP. In cAMP-dependent protein kinase (cAPK), this base has been proposed to be Asp 166, which corresponds to Asp 946 of the PI3K γ -946-DRH-948 sequence that is conserved in all PI3Ks. In the structure of PI3K γ , however, Asp 946 is not in a location where it could function as a general base catalyst. The constellation of residues in the active site in the presence of ATP–metal suggests that Asp 946 may simply have a structural role in maintaining the ATP-binding pocket. Therefore, either PI3K has no general base catalyst, in which case the mechanism could be primarily dissociative, involving a metaphosphate transition state¹², or a different residue assumes this function. One possible candidate is His 948; although its side chain is not near the γ -phosphate of ATP, a rotation around $\chi 1$ would place it in a location such that it could interact with the 3-hydroxyl of the lipid headgroup.

PI3Ks are one of the effectors for Ras proteins (reviewed in ref. 13). Binding of PI3K to Ras is affected by mutations in both switch I and switch II regions of Ras (residues 30–38 and 60–76, respectively)^{14,15}. These two regions change conformation upon GTP binding and are binding sites for a diverse array of downstream effectors; however, mutations in these switch regions differentially affect the binding of various effectors.

The RBD of PI3K γ (residues 220–311) has the same fold as the RBD of Raf¹⁶ and RalGDS¹⁷, two other well-characterized effectors of Ras (Fig. 5). The RBD of PI3K consists of a five-stranded mixed β -sheet (R $\beta 1$ –R $\beta 5$) flanked by two α -helices (R $\alpha 1$ and R $\alpha 2$). Residues 228–230 (in the R $\beta 1$ –R $\beta 2$ loop) and 257–265 (in the R $\alpha 1$ –R $\beta 3$ loop) are disordered.

The crystal structure of Ras-related protein Rap1A in complex with the RBD of protein kinase c-Raf¹⁶ and the structure of Ras in complex with the RBD of RalGDS¹⁷ suggest a structural basis for effector specificity. The structures of both of these complexes were determined using the isolated RBD, without the catalytic portions

Table 1 Data collection, structure determination and refinement statistics

Data collection and multiple isomorphous replacement phasing statistics

Data set	Resolution (Å)	Observations/unique reflections	Completeness (last shell) (%)	$R_{\text{merge}}^{\dagger\dagger}$	$\langle I/\sigma \rangle$ (last shell)	No. of sites	Phasing power	$R_{\text{iso}}^{\$ \$}$
Native ^{†††}	2.4	144,973/37,485	97.2 (90.6)	8.5	16.0 (3.1)	—	—	—
LuCl ₃ -1 ^{††}	2.2	191,292/49,599	95.5 (93.3)	9.5	14.3 (1.1)	7	1.7	0.23
LuCl ₃ -2 ^{††}	3.5	43,038/12,484	99.7 (98.2)	8.5	11.3 (3.4)	3	1.9	0.18
Lanthanides ^{§§}	3.0	71,426/19,180	97.9 (97.1)	4.5	15.6 (2.3)	8	1.9	0.24
ATM	2.7	94,900/25,688	92.6 (60.2)	4.8	17.0 (5.7)	5	0.8	0.22
Iodine ^{¶¶}	2.6	102,511/28,856	93.2 (67.1)	6.0	13.6 (1.4)	3	0.1	0.21

Refinement statistics

Data set	Resolution (Å)	Protein atoms	Waters	$R_{\text{crystal}}^{\text{ }}$	$R_{\text{free}}^{\text{ }}$ (% data)	R.m.s.d. from ideality ^{##}		
						Bonds	Angles	Dihedrals
LuCl ₃ -1	25.0–2.2	6,813	89	0.25	0.30 (5.4)	0.013 Å	1.7°	23°
Iodine ^{¶¶}	25.0–2.6	6,954	14	0.26	0.33 (5.0)	0.005 Å	1.1°	21°
Mn [#]	25.0–2.6	6,837	26	0.26	0.32 (5.6)	0.005 Å	1.2°	21°

Overall figure of merit 0.45

* The native crystal was soaked in 2.5 mM InsP₃, 1.0 mM ATP and 10 mM MgCl₂ for 1 h. Although this was the native crystal for heavy-atom phasing, the final high-resolution structure refinement used data from LuCl₃-1.

† LuCl₃-1 crystal was soaked in 20 mM LuCl₃ and 1.25 mM ATP for 1 h 40 min.

†† LuCl₃-2 crystal was soaked in 20 mM LuCl₃ and 1.3 mM ATP for 4 h.

§ Lanthanides crystal was soaked for 4 h in a mixture of 3.3 mM each of GdCl₃, TbCl₃, HoCl₃, ErCl₃, TmCl₃, and LuCl₃ with 1.26 mM ATP and 1 mM EMTS.

|| ATM crystal was soaked for 22 h in 10 mM sodium aurothiomalate.

¶ Iodine crystal was soaked for 75 min in 1 mM NaI₃ and 1 mM chloramine T. This crystal was originally prepared in an attempt to iodinate tyrosine residues as a heavy atom derivative, but no evidence of tyrosine iodination was seen in the resulting structure.

Mn crystal contained 1.4 mM ATP and 14 mM MnCl₂.

* Data were collected at ESRF beamline ID2b.

†† Data were collected at ESRF beamline ID14-4.

††† $R_{\text{merge}} = \sum_i \sum_j |I_i(hkl) - \langle I_i(hkl) \rangle| / \sum_i \sum_j I_i(hkl)$.

§§ $R_{\text{iso}} = \sum_i |F_{\text{deriv}}| - |F_{\text{native}}| / \sum_i |F_{\text{deriv}}|$.

||| The phasing power is defined as the ratio of the r.m.s. value of the heavy atom structure factor amplitudes and the r.m.s. value of the lack-of-closure error.

¶¶ R_{crystal} and $R_{\text{free}} = \sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i |F_{\text{obs}}|$; R_{free} calculated with the percentage of the data shown in parentheses.

R.m.s. deviations for bond angles and lengths in regard to Engh and Huber parameters.

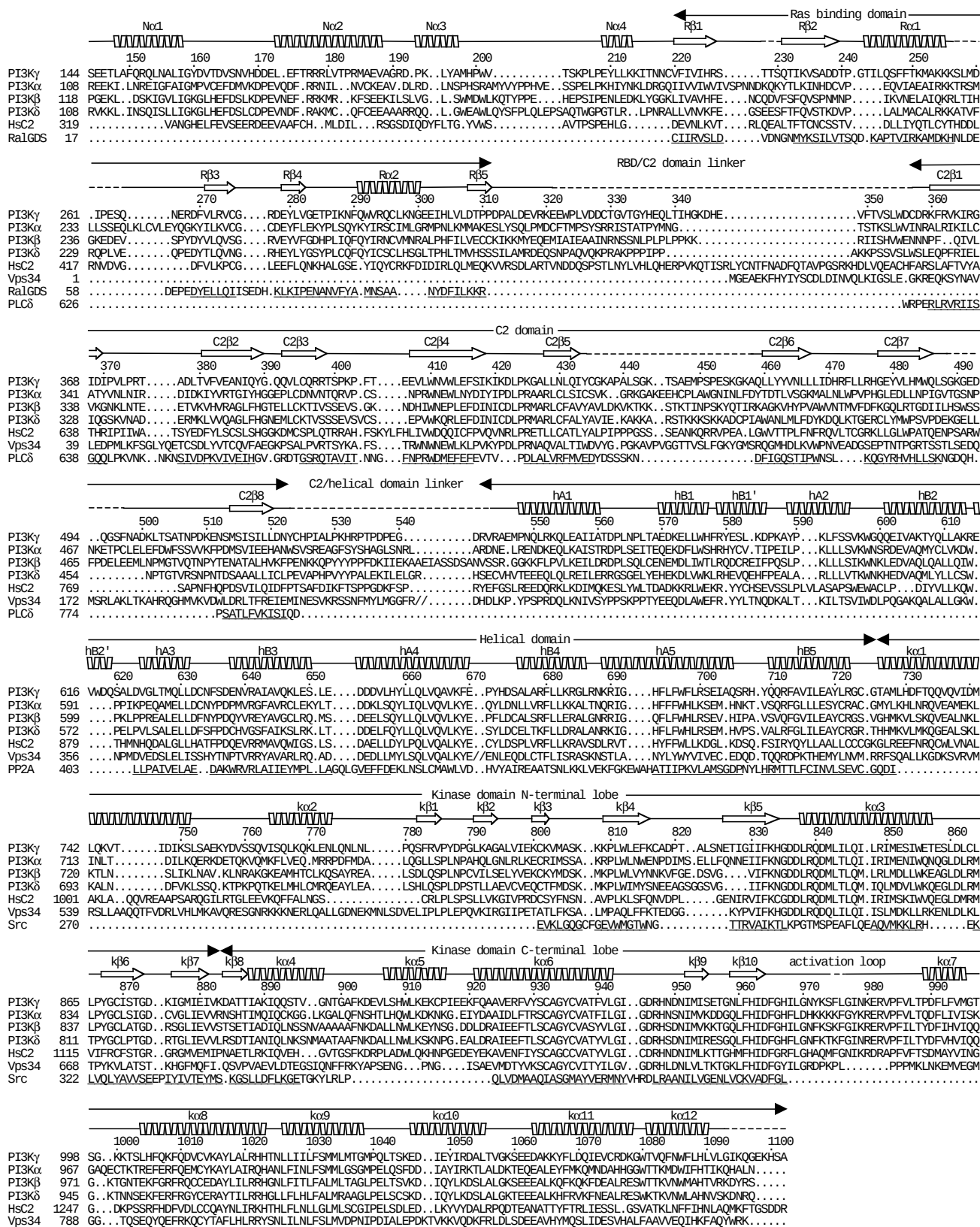


Figure 3 Secondary structure of the PI3K-γ p110 subunit and sequence alignments with other PI3K family members. Alignments are based on the Pfam database. The last line of the alignment illustrates proteins with structurally similar domains. Residues that superimpose with PI3Kγ are underlined. Sequences are porcine PI3Kγ (o02697), human PI3Kα (p42336), human PI3Kβ (p42338), human PI3Kδ (p00329), human class II PI3K

(HsC2-PI3K, o00750), human Vps34 (s57219), RalGDS (1lf), phospholipase Cδ (1dix), PR65A regulatory subunit (1b3u) and tyrosine kinase c-Src (1fmk). Dashed lines in the secondary structure indicate disordered residues and // indicates a large gap in the sequence.

of the effector molecules. The PI3K γ structure shows how the RBD interacts with the rest of the enzyme. The RBD of PI3K γ contacts the N-lobe and to a lesser degree the C-lobe of the catalytic domain. Residues in R α 2 and the R β 3–R β 4 loop interact with the catalytic domain, mainly with the k β 1–k β 2 and k β 4–k β 5 loops and helix k α 6. The relationship of the RBD to the rest of the enzyme suggests two possible mechanisms by which Ras binding may cause effector activation: a recruitment mechanism, in which Ras increases PI3K activity by translocating the enzyme to the plasma membrane; and an allosteric mechanism, in which Ras binding to the RBD causes a conformational change that would be propagated through the RBD/catalytic domain interface to affect substrate or cofactor binding.

By superimposing the RBDs of RalGDS and PI3K γ , we constructed a model of Ras interaction with PI3K γ (Fig. 5) from which we can rationalize the differential effects of various switch I and switch II mutants on PI3K binding compared with other effectors. T35S and D38E mutations in Ras switch I eliminate PI3K binding, but do not affect Raf binding¹⁵. The E37G mutation abolishes binding to PI3K and Raf but not to RalGDS. The Y40C mutation does not affect PI3K binding, but abrogates Raf and RalGDS binding. In the switch II region, the Y64G mutation eliminates PI3K and neurofibromin binding but has no effect on Raf binding¹⁴.

In our model of the PI3K–Ras interaction, residues Glu 37, Asp 38, Tyr 40 and Tyr 64 are at the PI3K–Ras interface. Lys 234 of PI3K could form a salt bridge to Glu 37 of Ras and Lys 255 at the C-

terminal end of R α 1 could form a salt bridge with Asp 38 of Ras. Lys 255 at the C-terminal end of R α 1 could form a salt bridge with Asp 38 of Ras. Lys 255 in PI3K γ is probably analogous to Lys 227 in PI3K α , in which mutation K227E blocks PI3K α binding to Ras¹⁵. Tyr 40 interacts with Lys 32 in RalGDS (numbering as in ref. 17); however, the very different orientation of the Lys 32 equivalent in PI3K γ (Lys 234) may not allow this interaction, which may account for the insensitivity of PI3K to the Y40C mutation. On the other hand, Tyr 64 in switch II is in a position to form a hydrogen bond with PI3K Asp 238, but this residue has no specific interaction with RalGDS, which may explain the sensitivity of PI3K to the Ras Y64G mutation.

The PI3K γ C2 domain (residues 357–522) is an eight-stranded antiparallel β -sandwich consisting of two four-stranded β -sheets (Fig. 6). The fold of this domain is the same as the type II C2 domain found in PLC δ 1¹⁸. The N-terminal regions of all three PI3K classes have C2 domains, whereas the class II enzymes have an additional C2 domain at the C terminus (Fig. 1). The segments leading from the RBD into the C2 domain and from the C2 domain to the helical domain are not ordered.

C2 domains are often involved in Ca²⁺-dependent or Ca²⁺-independent phospholipid membrane binding using three loops known as CBRs located at one end of the domain. The CBRs for PI3K γ are the loops connecting β 1 with β 2 (CBR1), β 3 with β 4 (CBR2), and β 5 with β 6 (CBR3). The CBR3 of PI3K γ is quite long compared with other C2 domains and is disordered in our structure.

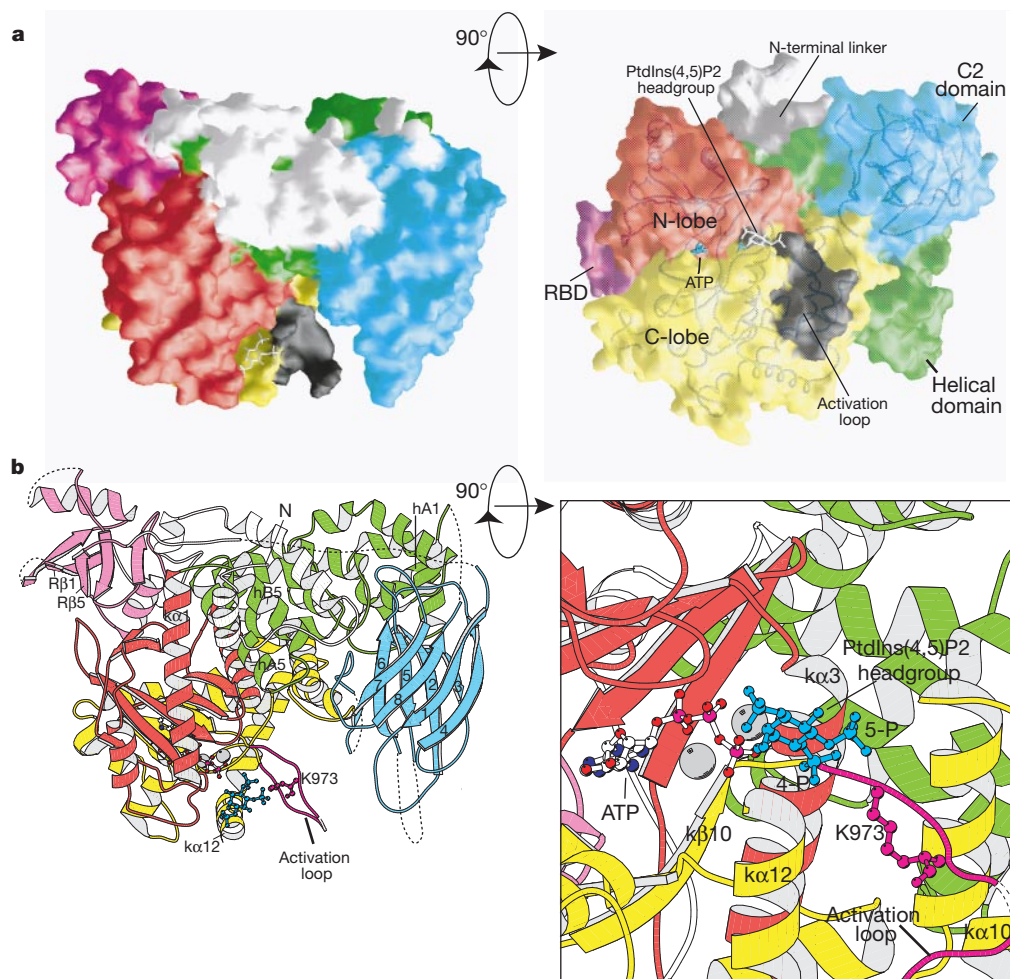
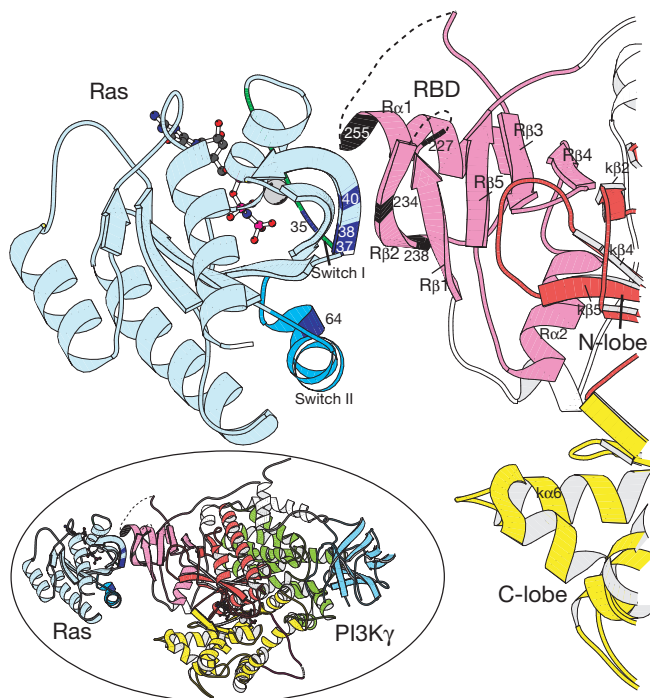


Figure 4 Model of phospholipid headgroup interactions with PI3K γ . **a**, Orthogonal views of the solvent-accessible surface. The activation loop is black. An inositol 1,4,5-trisphosphate (InsP₃) molecule (white ball-and-stick) is modelled in the active site with the

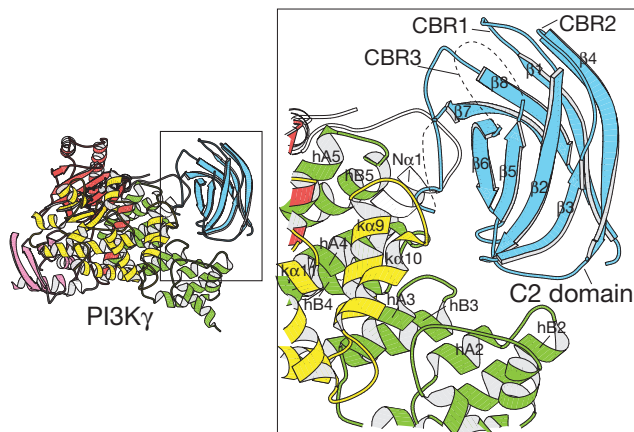
3-OH near the γ -phosphate of the bound ATP. **b**, The same views in ribbon representation showing the activation loop (magenta) and InsP₃ (blue). The right panel is expanded to show features of the putative headgroup interaction.

PI3K can bind phospholipid membranes in the absence of other protein components in a Ca^{2+} -independent manner and can carry out processive catalysis at the membrane surface. By analogy with enzymes such as protein kinase C and cytosolic phospholipase A2, the C2 domain of PI3K may participate in membrane interaction. Consistent with this, we found that the isolated C2 domain from PI3K γ binds multilamellar phospholipid vesicles similarly to the full-length enzyme (data not shown). In PI3K β and PI3K δ , CBR3

The structure of a type II β phosphatidylinositol phosphate kinase (PIPK) has been reported¹⁹. This dimeric enzyme, which phosphorylates phosphoinositides at the 4-hydroxyl, consists of a single, catalytic domain. The dimer has an extensive flat, positively charged surface which was proposed to be the membrane-binding interface of the enzyme. Although the N-lobe of PIPK is structurally related to the catalytic domain of PI3K γ , the location of PI3K γ C2 domain with respect to the catalytic domain sterically precludes membrane interactions using the surface of PI3K γ analogous to the putative PIPK membrane-binding surface. Given the location of the membrane-binding loops from the C2 domain and the cavity in the catalytic domain that must accommodate the PtdIns(4,5)P₂



blue stripes, whereas residues in the RBD of PI3K (purple) that are likely to be involved in Ras binding are shown as black stripes. Note the proximity of the RBD to the two lobes of the catalytic domain.



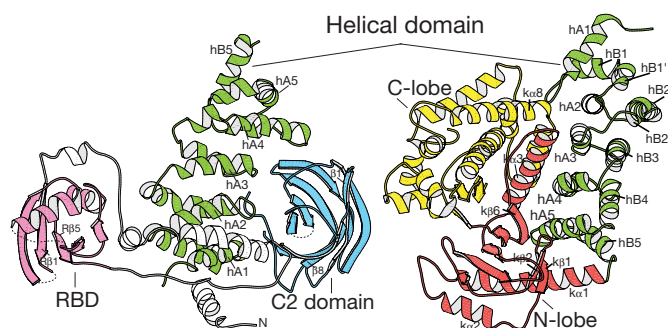


Figure 7 The helical domain (for colours see Fig. 1). The A/B antiparallel helical pairs characteristic of the HEAT motif topology consist of hA1/hB1, hA2/hB2, hA3/hB3, hA4/hB4 and hA5/hB5. Left, interaction between the helical domain and the RBD and C2 domain (the rest of the protein is removed for clarity). This interaction involves principally the A-helix surfaces. Right, interactions between the helical domain and the catalytic domain.

headgroup, the membrane-binding surface of PI3K γ could consist of the CBRs, the crevice between the N- and C-lobes of the catalytic domain, and the tip of the activation loop (Fig. 4a, right panel represents a view from the membrane surface).

The helical domain of PI3K (545–725) consists of five A/B pairs of anti-parallel helices (Fig. 7). The first two pairs have one kinked helix each, hB1/hB1' and hB2/hB2'. This region has been variously referred to as HR2, the PI3K accessory domain and the PIK domain, but its function is not known. The paired arrangement of a series of helices connected into a right-handed superhelix is reminiscent of the PR65/A regulatory subunit of protein phosphatase 2A (PP2A)²⁰. PR65/A is a member of a diverse group of proteins that contain 3–25 tandem repeats of a short sequence, called the HEAT motif. The HEAT motif consists of paired helices A and B arranged so that the A and B helices within a pair are antiparallel, and the A and B helices from one motif are parallel to the A and B helices of the next motif in the sequence. Although no HEAT sequence motif is apparent in the helical domain of PI3K, its structure is quite similar to that of PR65/A in terms of the arrangement of helices, the length of the A/B units, and the angle between the A/B pairs.

The function of HEAT repeats is to form protein–protein interactions. In importin- β , interactions with the small GTPase Ran involve the surfaces of the B helices²¹. In PR65/A, mutagenesis has implicated the loops connecting the A/B pairs as the region responsible for interaction with PP2A²⁰. In PI3K γ , the helical domain is central to the interdomain packing: the surface formed by the A helices interacts with the catalytic domain; the loops connecting A and B helices within a pair pack against the C2 domain; and the loops between helical pairs pack against the RBD (Fig. 7). Much of the 'B' surface is solvent exposed and may interact with other proteins that bind PI3K γ , such as the p101 adaptor or G β subunits.

The helical domain is common to both PI3K and PI4K families and serves as a spine on which the other domains are fastened. One of the proteins in which the HEAT sequence motif was first noted is the target of rapamycin, TOR, a yeast homologue of human FRAP (reviewed in ref. 22). FRAP has a C-terminal domain with clear sequence homology to the catalytic domain of PI3Ks. The secondary-structure prediction for the remainder of FRAP suggests that most of FRAP, apart from the catalytic domain, may consist of helical repeats folded into a right-handed superhelix as in the helical domain of PI3K γ .

This first structure of a PI3K provides a framework within which mutagenesis and detailed kinetic studies can be carried out to establish the enzymatic mechanism and the mode of activation by Ras and heterotrimeric G-protein subunits. □

Methods

Protein expression, purification and crystallization

The cell-free extract of baculovirus-infected Sf9 cells expressing the His-tagged catalytic subunit of porcine PI3K γ (residues 1–143 deleted) was used for protein purification using Talon resin, followed by thrombin cleavage, anion and cation exchange, and gel filtration chromatography. Crystals were grown by mixing 1 μ l of PI3K (3.5–4.0 mg ml⁻¹, in a buffer containing 20 mM Tris-HCl pH7.2, 1% v/v ethylene glycol, 1% w/v betaine, 0.02% w/v CHAPS and 5 mM dithiothreitol) with 1 μ l of a reservoir solution containing 150–200 mM Li₂SO₄, 100 mM Tris-HCl pH 7.25 and 14–15% PEG 4000.

Data collection and structure determination

Crystals have C2 symmetry with unit-cell dimensions of $a = 143.3$ Å, $b = 67.6$ Å, $c = 107.0$ Å, $\beta = 95.9^\circ$, and contain one protein molecule in the asymmetric unit. Diffraction data were collected at ESRF beamlines ID2 and ID14-4 at 100K after freezing crystals in a cryoprotectant consisting of 150–200 mM Li₂SO₄, 100 mM Tris-HCl pH 7.25, 12% glycerol and 20% PEG 4000. Data were processed using MOSFLM²³ and CCP4 programs²⁴. The structure was determined by multiple isomorphous replacement methods. Heavy-atom positions were located using Solve²⁵ and refined with Sharp²⁶ (Table 1). A model was built into the electron-density maps using the program O²⁷ and refined using CNS²⁸. The average B-factor for all atoms is 60 Å². The structure has no residues in disallowed regions of the Ramachandran plot.

The highest resolution data obtained were for the complex containing ATP and lutein; refinement resulted in a model with a free R-factor of 0.30 to a resolution of 2.2 Å. This complex has 854 residues visible in the electron-density map. Crystals with and without ATP had only minor differences in side-chain conformations in the active-site residues. PI3Ks require a Mg²⁺ or Mn²⁺ cofactor for enzymatic activity. In complexes with Lu³⁺, Mg²⁺ or Mn²⁺, each of the metals binds at the same two sites.

Received 10 August; accepted 6 October 1999.

1. Toker, A. & Cantley, L. C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **387**, 673–676 (1997).
2. Domin, J. & Waterfield, M. D. Using structure to define the function of phosphoinositide 3-kinase family members. *FEBS Lett.* **410**, 91–95 (1997).
3. Stoyanov, B. et al. Cloning and characterisation of a G protein-activated human phosphoinositide-3 kinase. *Science* **269**, 690–693 (1995).
4. Stephens, L. R. et al. The G β sensitivity of a PI3K is dependent upon a tightly associated adaptor, p101. *Cell* **89**, 105–114 (1997).
5. Krugmann, S., Hawkins, P. T., Pryer, N. & Braselmann, S. Characterizing the interactions between the two subunits of the p101/p110 γ phosphoinositide 3-kinase and their role in the activation of this enzyme by G β subunits. *J. Biol. Chem.* **274**, 17152–17158 (1999).
6. Taylor, S. S. et al. Catalytic subunit of cyclic AMP-dependent protein kinase: structure and dynamics of the active site cleft. *Pharmacol. Ther.* **82**, 133–141 (1999).
7. Dhand, R. et al. PI3-kinase is a dual specificity enzyme—autoregulation by an intrinsic protein serine kinase activity. *EMBO J.* **13**, 522–533 (1994).
8. Stack, J. H. & Emr, S. D. Vps34p required for yeast vacuolar protein sorting is a multiple specificity kinase that exhibits both protein kinase and phosphatidylinositol-specific PI-3-kinase activities. *J. Biol. Chem.* **269**, 31552–31562 (1994).
9. Wymann, M. P. et al. Wortmannin inactivates phosphoinositide 3-kinase by covalent modification of Lys-802, a residue involved in the phosphate transfer reaction. *Mol. Cell. Biol.* **16**, 1722–1733 (1996).
10. Bondeva, T. et al. Bifurcation of lipid and protein kinase signals of PI3K γ to the protein kinases PKB and MAPK. *Science* **282**, 293–296 (1998).
11. Vanhaesebroeck, B. et al. Autophosphorylation of p110 δ phosphoinositide 3-kinase: a new paradigm for the regulation of lipid kinases *in vitro* and *in vivo*. *EMBO J.* **18**, 1292–1302 (1999).
12. Kim, K. & Cole, P. A. Measurement of a Brønsted nucleophile coefficient and insights into the transition state for a protein tyrosine kinase. *J. Am. Chem. Soc.* **119**, 11096–11097 (1997).
13. Marshall, C. J. Ras effectors. *Curr. Opin. Cell Biol.* **8**, 197–204 (1996).
14. Moodie, S. A. et al. Different structural requirements within the switch II region of the Ras protein for interactions with specific downstream targets. *Oncogene* **11**, 447–454 (1995).
15. Rodriguez-Viciana, P. et al. Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. *Cell* **89**, 457–467 (1997).
16. Nassar, M. P. et al. The 2.2 Å crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature* **375**, 554–560 (1995).
17. Huang, L., Hofer, F., Martin, G. S. & Kim, S.-H. Structural basis for the interaction of Ras with RalGDS. *Nature Struct. Biol.* **5**, 422–426 (1998).
18. Essen, L.-O., Perisic, O., Lynch, D. E., Katan, M. & Williams, R. L. A ternary metal binding site in the C2 domain of phosphoinositide-specific phospholipase C- δ 1. *Biochemistry* **36**, 2753–2762 (1997).
19. Rao, V. D., Misra, S., Boronnikov, I. V., Anderson, R. A. & Hurley, J. H. Structure of type II β phosphatidylinositol phosphate kinase: a protein kinase fold flattened for interfacial phosphorylation. *Cell* **94**, 829–839 (1998).
20. Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Barford, D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96**, 99–110 (1999).
21. Vetter, I. R., Arndt, A., Kutay, U., Görlich, D. & Wittinghofer, A. Structural view of the Ran-importin β interaction at 2.3 Å resolution. *Cell* **97**, 635–646 (1999).
22. Dennis, P. B., Fumagalli, S. & Thomas, G. Target of rapamycin (TOR): balancing the opposing forces of protein synthesis and degradation. *Curr. Opin. Genet. Dev.* **9**, 49–54 (1999).
23. Leslie, A. G. W. Recent changes to the MOSFLM package for film and image plate data, in *Joint CCP4 and ESRF-EACMB Newsletter on Protein Crystallography* Vol. 26 (Daresbury Laboratory, Warrington, UK, 1992).
24. CCP4. Collaborative Computing Project 4: A suite of programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
25. Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).

26. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
27. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
29. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
30. Yamaguchi, H. & Hendrickson, W. A. Structural basis for the activation of human lymphocyte kinase Lck tyrosine phosphorylation. *Nature* **384**, 484–489 (1996).
31. Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

Acknowledgements

We thank D. Lynch for technical assistance, S. Krugmann for PI3K assays, the staff of synchrotron beamlines ID2b, ID14-3 and ID14-4 at ESRF, IMCA-CAT at APS, X11 at EMBL Hamburg, Station 9.6 at Daresbury SRS, and Elettra, Italy, for help in synchrotron data collection. We thank P. Roversi and G. Bricogne for their assistance with BUSTER, R. Rubin and C. Humblet for access to the IMCA-CAT beamline, and P. Hawkins and A. Murzin for helpful discussions. C.R. was supported by the Deutsche Forschungsgemeinschaft. We are grateful for support from the BBSRC (via a senior fellowship to P. Hawkins), the British Heart Foundation, Parke-Davis and Onyx Pharmaceuticals (R.L.W.).

Correspondence and requests for materials should be addressed to R.L.W. (e-mail: rlw@mrclimb.cam.ac.uk). The accession code for the atomic coordinates of the structure at the Protein Data Bank is 1qmm.

Sustained oscillations in living cells

Sune Danø, Preben Graae Sørensen & Finn Hynne

Department of Chemistry and CATS, H.C. Ørsted Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark

Glycolytic oscillations in yeast have been studied for many years simply by adding a glucose pulse to a suspension of cells and measuring the resulting transient oscillations of NADH^{1–12}. Here we show, using a suspension of yeast cells, that living cells can be kept in a well defined oscillating state indefinitely when starved cells, glucose and cyanide are pumped into a cuvette with outflow of surplus liquid. Our results show that the transitions between stationary and oscillatory behaviour are uniquely described mathematically by the Hopf bifurcation¹³. This result characterizes the dynamical properties close to the transition point. Our perturbation experiments show that the cells remain strongly coupled very close to the transition. Therefore, the transition takes place in each of the cells and is not a desynchronization phenomenon. With these two observations, a study of the kinetic details of glycolysis, as it actually takes place in a living cell, is possible using experiments designed in the framework of non-linear dynamics. Acetaldehyde is known to synchronize the oscillations¹⁰. Our results show that glucose is another messenger substance, as long as the glucose transporter is not saturated.

The ubiquitous glycolytic pathway is the first step in sugar catabolism, producing ATP, NADH and pyruvate. Under anaerobic conditions, the NADH is reused in the subsequent fermentation of the pyruvate. Transient glycolytic oscillations are known to occur in suspensions of yeast cells^{1,2}; see ref. 12 for measurements of most of the relevant metabolites during oscillatory glycolysis. The observed bulk oscillations of intracellular NADH depend on the ability of individual cells to synchronize their oscillations. This capability was proven by mixing two suspensions oscillating 180° out of phase: the bulk oscillations of NADH disappear immediately after the mixing, but reappear spontaneously after some time^{3,4,10}.

Synchronized bulk oscillations depend on a sufficiently high cell density⁶. They are promoted by anaerobiosis, especially when induced by cyanide^{4,9}, which is a potent inhibitor of cytochrome *c* oxidase in the respiratory chain.

Glycolytic oscillations are normally induced by a pulse of glucose, and the resulting oscillations change gradually as the excess of extracellular glucose is used up. An attempt was made to extend the duration of the oscillations using an infusion system where glucose was slowly injected into the cell suspension¹⁴. However, owing to cell ageing, accumulation of waste products or dilution of the suspension, the observed oscillations were still transient. In both the infusion experiment and the glucose-pulse experiment, the duration of the oscillating transient depends on the enzymatic composition of the yeast cells^{3,7,8,14}. In the glucose-pulse experiments, the longest transient trains of oscillations are found when a saturated, high-affinity glucose transporter provides the cell with an almost constant flow of glucose throughout the transient¹¹. In our experimental setup, a continuous-flow stirred tank reactor (CSTR), we control the glycolytic flow by means of the inflows, to give truly sustained oscillations.

To obtain these oscillations, we prepare *Saccharomyces cerevisiae* as described in ref. 8. The yeast is grown in batch culture at 30 °C to the point of glucose depletion. The cells are then removed from the growth medium and starved for a couple of hours, before being resuspended in phosphate buffer and kept below 5 °C. The resulting suspension flows into a stirred and thermostat-regulated reactor through a peristaltic pump. Solutions of glucose and cyanide flow into the reactor through two stepper-motor controlled piston burettes (Fig. 1). We monitor the oscillations by measuring NADH fluorescence.

Figure 2 shows a typical recording of stable undamped oscillations in the cell suspension. In principle the oscillations could continue forever with constant amplitude and period. This is possible because our setup is a truly open system. In practice, the amount of yeast cells available from the batch growth limits the duration of the run to 14 h in our experiments.

The environmental conditions of the cells can be changed by altering the flow rates of the inflows. Depending on flow rate we observe either a stationary state or continuous oscillations. The flow rate where the stationary state becomes unstable is known as the bifurcation point. As the flow rate is changed from this critical value, the amplitude of the oscillations grows in proportion to the square root of the deviation from the critical value. For a dynamic system such behaviour is characteristic of a supercritical Hopf bifurcation (SHB). This description is, of course, only valid sufficiently close to the bifurcation point. In Fig. 3 we show how the amplitude varies with the glucose flow rate. A similar agreement with bifurcation theory is seen when the flow rate of cyanide is

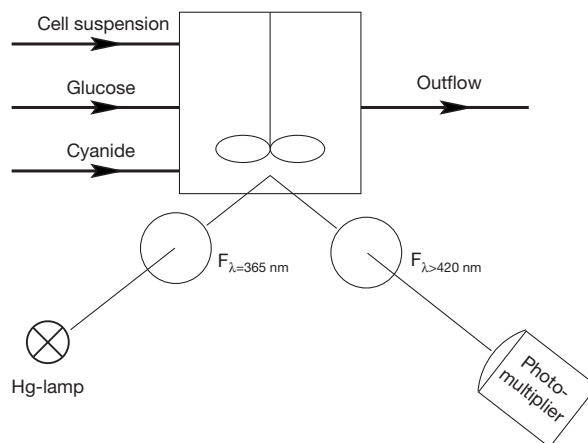


Figure 1 The experimental setup. F designates optical filters. See details in the Methods section.

changed, but the cyanide flow has both a lower and an upper SHB. The suppression of oscillations by excess cyanide can be explained by a reaction between cyanide and acetaldehyde⁹. On the oscillatory side of an SHB the system has an almost elliptical stable limit cycle with an unstable stationary point at the centre. Figure 2 shows sinusoidal oscillations in the concentration of NADH as the state of the system goes round the limit cycle.

If one of the extracellular concentrations is changed by instantaneous addition of a small amount of the corresponding species, the state of the system will show some transient behaviour and will eventually return to the limit cycle. A characteristic response which we call quenching is obtained if the concentration of an extracellular species is perturbed so that the transient passes close to the unstable stationary point. Here the concentrations are almost constant and, after the passage, the system makes a spiralling return to the limit cycle. Working at a stable focus, Winfree has demonstrated the phase-resetting effect of O₂ in a cyanide-free system, and related the experiments to the geometry of concentration space⁵. The relationship between quenching and phase-resetting experiments is described in ref. 15.

The quenching behaviour is demonstrated as follows. By adjusting the flows of glucose and cyanide, an operating point with small sinusoidal oscillations is selected. At some chosen phase a certain amount of the substance of interest is instantaneously added to the suspension in the reactor. We find (initially by trial and error) that it is possible to obtain the quenching response by adding a definite amount of either glucose or acetaldehyde at a particular phase of oscillation (characteristic for each substance). For all other phases and amounts of addition the response will be different (Fig. 4).

Attempts to quench with cyanide, ethanol or pyruvate were unsuccessful. Pyruvate induces a slow, phase-independent response, whereas ethanol produces no response at all. Cyanide perturbations increase the NADH level, and can diminish the amplitude of the oscillations significantly, but cyanide cannot quench the oscillations because it reacts too slowly with the rest of the system.

Quenching with acetaldehyde is possible for a wide range of flow rates, but quenching with glucose is only possible for low glucose flow rates. The upper limit of glucose flow for which quenching with glucose is possible coincides with the sudden stabilization of the amplitude at higher flow rates (Fig. 3). We interpret this as representing saturation of the glucose transporter.

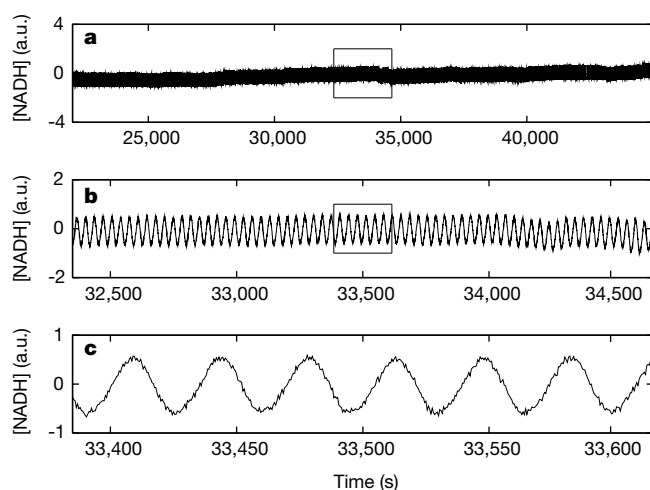


Figure 2 Sustained oscillations in *Saccharomyces cerevisiae*. **a**, An unperturbed part of a CSTR experiment. The period of the oscillations is 34.8 ± 0.20 s, and the amplitude is constant with a relative standard deviation of 2.8%. **b**, An expansion of the box in the top panel. **c**, A similar expansion of the box in the middle panel. The cell density is 1.61×10^9 cells per ml and the specific flow rate is $50.6 \times 10^{-3} \text{ min}^{-1}$. The mixed flow concentrations of glucose and cyanide are 35.0 mM and 5.38 mM. The experiment starts at time zero. a.u., arbitrary units.

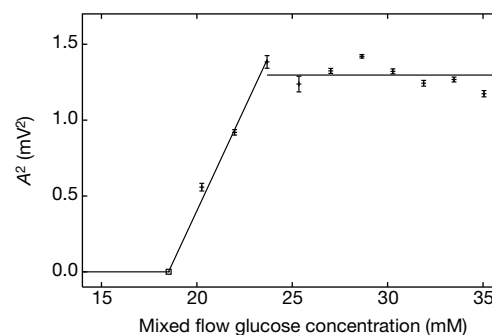


Figure 3 Plot of the square of the amplitude A as a function of glucose flow rate. The bifurcation point is found at a mixed flow glucose concentration of 18.5 mM. Oscillations were sinusoidal throughout the whole range of measurements. Cell density, mixed flow cyanide concentration and specific flow rate were fixed at 1.61×10^9 cells per ml, 5.60 mM and $4.86 \times 10^{-3} \text{ min}^{-1}$, respectively.

The experimental results summarized in Figs 2–4 clearly show that the bifurcation found in the yeast cells is an SHB. Hopf bifurcations are ubiquitous in physical and chemical systems but have never before been demonstrated in a living system. Its presence creates new possibilities for investigating *in vivo* metabolic dynamics, as quenching experiments with known amounts of all relevant substances allow the underlying kinetics of the system to be revealed and described quantitatively^{16–19}.

The possible uses of the quenching method for investigating a suspension of yeast cells are far from exhausted by these experiments, but our results already contain valuable biological information. They strongly support a previous suggestion that acetaldehyde is a synchronizing agent in yeast¹⁰, but at the same time they show that no single substance can be designated as ‘the synchronizer’, as glucose can act in the same way when the glucose transporter is not saturated. The experimental results may also be used to test and improve comprehensive models of yeast cell suspensions (S.D., P.G.S. and F.H., manuscript in preparation).

The fact that macroscopic oscillations are seen indicates synchronization of the oscillations in the cells. Models of interacting oscillators²⁰ have shown that this synchronization could be strong, with all yeast cells oscillating almost exactly in phase, or it could be weak, with a broad distribution of phases. Similarly, the disappearance of the oscillations at the critical point could result either from a gradual widening of the phase profile or from an SHB in each of the strongly coupled oscillators. Our quenchings show that the yeast cells are in fact strongly coupled, as the phase interval in which a given perturbation will stop the oscillations is quite narrow ($<10^\circ$; compare Fig. 4 b and c). As glucose quenchings are only possible close to the SHB at low glucose flow rates, the phase distribution of the oscillators is still very narrow close to the bifurcation point. This means that the disappearance of the oscillations corresponds to an SHB in each of the yeast cells, as opposed to a desynchronization of the oscillators.

Detailed understanding of the oscillations in glycolysis and their synchronization may be relevant to other cell systems. Glycolytic oscillations have been reported in many other cell types, notably muscle cells^{21,22} pancreatic β -cells²³ and heart cells^{24,25}. The possible coupling of the glycolytic oscillator to other cellular processes through oscillations in the [ATP]/[ADP] ratio might be important for the pulsatory insulin secretion of β -cells^{26,27} or in the spatio-temporal regulation of blood flow²⁸. Furthermore, a synchronization mechanism, similar to the one seen in our experiments, could work through the blood as well. This could explain the existence of bulk oscillations of insulin concentration in the blood.

Quenching of glycolytic oscillations with acetaldehyde shows the importance of the coupling of fermentation and glycolysis through

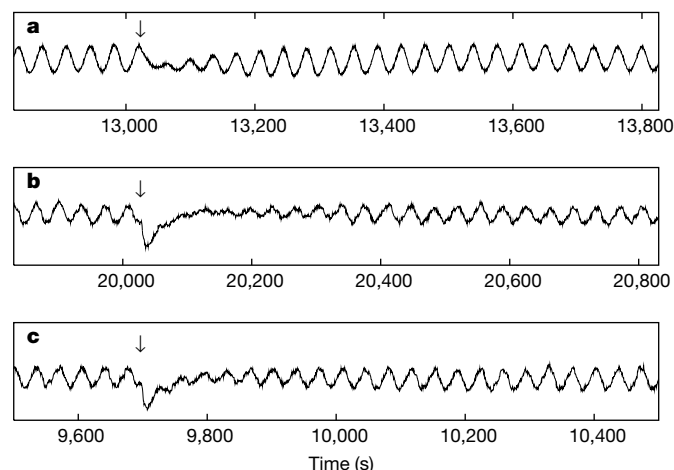


Figure 4 Fluorescence traces showing the response to instantaneous addition of extracellular glucose or acetaldehyde. **a**, Addition of glucose at a phase of 4° (arrow) with a change in glucose concentration of 1.11 mM, which quenches the oscillations. **b**, Quenching with 98 μ M acetaldehyde at 172° (arrow). The uniqueness of the quenching phase and concentration change is illustrated in **c**, which repeats the experiment in **b** exactly, except that the phase is 180° (arrow) instead of 172° : the amplitude is diminished, but the oscillations do not stop completely. For details, see Methods.

NAD⁺/NADH. We have not been able to quench the oscillations with pyruvate, even though pyruvate has a similar effect on the redox balance as acetaldehyde. Analogous to the saturation of the glucose transporter, this could be due to the low permeability of the cell membrane to pyruvate. Cell types that ferment pyruvate to lactic acid instead of ethanol might synchronize glycolytic oscillations by regulating either the pyruvate permeability of the membrane or the rate of glucose uptake.

Our experiments show that it is possible to control the exact dynamic state of a living system in a flow reactor. This applies not only to oscillatory states, but also to maintaining a true steady state. Several steady states can be compared, and each steady state can be investigated for extended periods of time. This allows new ways of performing biochemical investigations, such as time-averaged nuclear magnetic resonance studies of metabolism as it actually takes place in the living cell.

Yeast cells are basically unicellular organisms when in their natural environment. Nevertheless, the glycolytic clock provides a hidden potential for oscillatory communication. In our experiments the clock is awakened by the unnatural environmental conditions imposed by the special harvesting and addition of cyanide. From the theory of dynamical systems, it is known that a Hopf bifurcation is not a singular point in parameter space, but implies an infinity of Hopf points on a 'surface' possibly extending to conditions which might arise spontaneously in a changing natural environment. Thus, our experiments suggest an evolutionary path from unicellular to multicellular behaviour. □

Methods

The experimental setup

See Fig. 1. The reactor is a stirred 1 × 1 cm cuvette, maintained at 25 °C. The yeast suspension, which has a cell density of ~10% by weight, is kept cold and stirred to avoid sedimentation. From the reservoir it is pumped into the reactor by a peristaltic pump with a constant rate of 80.2 μ l min⁻¹. To avoid agglomeration of yeast cells in the tubes of the peristaltic pump, we inject bubbles of air into the tubes at a constant rate. Stepper-motor driven piston burettes provide two additional small and adjustable flows of glucose and cyanide solutions. The volume is kept constant at 1.89 ± 0.11 ml, and the mean residence time is 23.6 min. NADH fluorescence is achieved by excitation with the 365 nm mercury line, and is measured as the emission of light in the interval 420–660 nm with a photomultiplier. Owing to the opacity of the cell suspension, the fluorescence originates

from a thin layer near the surface of the cuvette. To compensate for variations in the output of the mercury lamp the intensity is measured by a photodetector at 365 nm and used to correct the photomultiplier signal.

Perturbation experiments

In the perturbation experiment with extracellular glucose (Fig. 4a) the cell density is 1.62×10^9 cells per ml, and the mixed flow concentrations of glucose and cyanide are 23.1 mM and 4.26 mM, respectively. The specific flow rate is 50.6×10^{-3} min⁻¹. In the perturbation experiments with extracellular acetaldehyde (Fig. 4b, c), the cell density is 1.66×10^9 cells per ml and the mixed flow concentrations of glucose and cyanide are 35.0 mM and 5.37 mM, respectively. The specific flow rate is 47.9×10^{-3} min⁻¹.

Received 23 April; accepted 21 September 1999.

- Duysens, L. N. M. & Ames, J. Fluorescence spectrophotometry of reduced phosphopyridine nucleotides in intact cells in the near-ultraviolet and visible region. *Biochim. Biophys. Acta* **24**, 19–26 (1957).
- Ghosh, A. & Chance, B. Oscillations of glycolytic intermediates in yeast cells. *Biochem. Biophys. Res. Commun.* **16**, 174–181 (1964).
- Pye, E. K. Biochemical mechanisms underlying the metabolic oscillations in yeast. *Can. J. Bot.* **47**, 271–285 (1969).
- Ghosh, A. K., Chance, B. & Pye, E. K. Metabolic coupling and synchronization of NADH oscillations in yeast cell populations. *Arch. Biochem. Biophys.* **145**, 319–331 (1971).
- Winfree, A. Oscillatory glycolysis in yeast: the pattern of phase resetting by oxygen. *Arch. Biochem. Biophys.* **149**, 388–401 (1972).
- Aldridge, J. & Pye, E. K. Cell density dependence of oscillatory metabolism. *Nature* **259**, 670–671 (1976).
- Kreuzberg, K. H. & Betz, A. Amplitude and period length of yeast NADH oscillations fermenting on different sugars in dependence of growth phase, starvation and hexose concentration. *J. Interdiscipl. Cycle Res.* **10**, 41–50 (1979).
- Richard, P., Teusink, B., Westerhoff, H. V. & van Dam, K. Around the growth phase transition *S. cerevisiae*'s make-up favours sustained oscillations of intracellular metabolites. *FEBS Lett.* **318**, 80–82 (1993).
- Richard, P. *et al.* Yeast cells with a specific cellular make-up and an environment that removes acetaldehyde are prone to sustained glycolytic oscillations. *FEBS Lett.* **341**, 223–226 (1994).
- Richard, P., Bakker, B. M., Teusink, B., van Dam, K. & Westerhoff, H. V. Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *Eur. J. Biochem.* **235**, 238–241 (1996).
- Teusink, B. *et al.* Synchronized heat flux oscillations in yeast cell populations. *J. Biol. Chem.* **271**, 24442–24448 (1996).
- Richard, P., Teusink, B., Hemker, B. B., van Dam, K. & Westerhoff, H. V. Sustained oscillations in free-energy state and hexose phosphates in yeast. *Yeast* **12**, 731–740 (1996).
- Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms* page 6 (Cambridge Univ. Press, Cambridge, 1996).
- Hess, B. & Boiteux, A. Mechanism of glycolytic oscillation in yeast, I. *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 1567–1574 (1968).
- Kosek, J., Sørensen, P. G., Marek, M. & Hynne, F. Normal form analysis of the Belousov-Zhabotinsky reaction close to a Hopf bifurcation. *J. Phys. Chem.* **98**, 6128–6135 (1994).
- Hynne, F., Sørensen, P. G. & Nielsen, K. Quenching of chemical oscillations: General theory. *J. Chem. Phys.* **92**, 1747–1757 (1990).
- Vukojević, V., Sørensen, P. G. & Hynne, F. Quenching analysis of the Briggs-Rauscher reaction. *J. Phys. Chem.* **97**, 4091–4100 (1993).
- Vukojević, V., Sørensen, P. G. & Hynne, F. Predictive value of a model of the Briggs-Rauscher reaction fitted to quenching experiments. *J. Phys. Chem.* **100**, 17175–17185 (1996).
- Hynne, F., Sørensen, P. G. & Møller, T. Complete optimization of models of the Belousov-Zhabotinsky reaction at a Hopf bifurcation. *J. Chem. Phys.* **98**, 219–230 (1993).
- Matthews, P. C., Mirollo, R. E. & Strogatz, S. H. Dynamics of a large system of coupled nonlinear oscillators. *Physica D* **52**, 293–331 (1991).
- Frenkel, R. Control of reduced diphosphopyridine nucleotide oscillations in beef heart extracts. II. Oscillations of glycolytic intermediates and adenine nucleotides. *Arch. Biochem. Biophys.* **125**, 157–165 (1968).
- Tornheim, K. & Lowenstein, J. M. The purine nucleotide cycle. III. Oscillations in metabolite concentrations during the operation of the cycle in muscle extracts. *J. Biol. Chem.* **248**, 2670–2677 (1973).
- Chou, H.-F., Berman, N. & Ipp, E. Oscillations of lactate released from islets of Langerhans: evidence for oscillatory glycolysis in β -cells. *Am. J. Physiol.* **262**, E800–E805 (1992).
- O'Rourke, B., Ramza, B. M. & Marban, E. Oscillations of membrane current and excitability driven by metabolic oscillations in heart cells. *Science* **265**, 962–966 (1994).
- O'Rourke, B., Ramza, B. M., Romashko, D. N. & Marban, E. Metabolic oscillations in heart cells. *Adv. Exp. Med. Biol.* **382**, 165–174 (1995).
- Tornheim, K. Are metabolic oscillations responsible for normal oscillatory insulin secretion? *Diabetes* **46**, 1375–1380 (1997).
- Civelek, V. N., Deeney, J. T., Fuson, G. E., Corkey, B. E. & Tornheim, K. Oscillations in oxygen consumption by permeabilized clonal pancreatic β -cells (HIT) incubated in an oscillatory glycolyzing muscle extract. Roles of free Ca^{2+} , substrates, and the ATP/ADP ratio. *Diabetes* **46**, 51–56 (1997).
- Siegel, G., Malmsten, M., Klüssendorf, D. & Hofer, H.-W. Vascular smooth muscle, a multiply feedback-coupled system of high versatility, modulation and cell-signaling variability. *Int. J. Microcirc.* **17**, 360–373 (1997).

Acknowledgements

We thank B. Teusink for advice on the growth and handling of the yeast cells.

Correspondence and requests for materials should be addressed to S.D. (e-mail: sdd@osc.kiku.dk).

A ribozyme that lacks cytidine

Jeff Rogers & Gerald F. Joyce

Departments of Chemistry and Molecular Biology, and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

The RNA-world hypothesis proposes that, before the advent of DNA and protein, life was based on RNA, with RNA serving as both the repository of genetic information and the chief agent of catalytic function¹. An argument against an RNA world is that the components of RNA lack the chemical diversity necessary to sustain life. Unlike proteins, which contain 20 different amino-acid subunits, nucleic acids are composed of only four subunits which have very similar chemical properties. Yet RNA is capable of a broad range of catalytic functions^{2–7}. Here we show that even three nucleic-acid subunits are sufficient to provide a substantial increase in the catalytic rate. Starting from a molecule that contained roughly equal proportions of all four nucleosides, we used *in vitro* evolution to obtain an RNA ligase ribozyme that lacks cytidine. This ribozyme folds into a defined structure and has a catalytic rate that is about 10⁵-fold faster than the uncatalysed rate of template-directed RNA ligation.

The decision to eliminate cytidine was based on two factors. First, cytidine is the least stable of the four nucleosides because

of its tendency to undergo spontaneous deamination to uridine ($t_{1/2} = 340$ yr at pH 7 and 25°C). For this reason, it has been suggested that cytidine was not present in the first genetic material⁸. Second, nucleic acids containing only adenosine, guanosine and uridine may still form both A-U Watson–Crick and G-U wobble pairs, which might be sufficient to form complex secondary and tertiary structures.

An RNA ligase ribozyme, called the class I ligase, has been obtained through *in vitro* evolution^{9,10}. This molecule contains roughly 140 nucleotides and forms 7 paired elements. It catalyses the joining of the 3'-hydroxyl of a template-bound oligonucleotide substrate to the 5'-triphosphate of the ribozyme. The ribozyme has been made to undergo continuous *in vitro* evolution by replacing the usual oligonucleotide substrate with a chimaeric DNA/RNA substrate that contains the sequence of the promoter element for T7 RNA polymerase¹¹. We chose the product of the first 52 h of continuous *in vitro* evolution, the E100 ribozyme, as the starting point for *in vitro* evolution of a cytidine-free ligase ribozyme.

A pool of 10¹⁴ variants was constructed by polymerase chain reaction (PCR) mutagenesis of DNA encoding the E100 ribozyme, followed by treatment of the DNA with sodium bisulphite. Sodium bisulphite greatly accelerates the deamination of cytosine to uracil¹², mimicking the pathway for spontaneous decomposition of cytosine. Cytosine deamination was performed at the DNA rather than the RNA level because this allowed it to be carried out under strongly denaturing conditions (elevated temperature and pH), thus preventing any cytidine residues from being sequestered within secondary and tertiary structure. Such harsh conditions

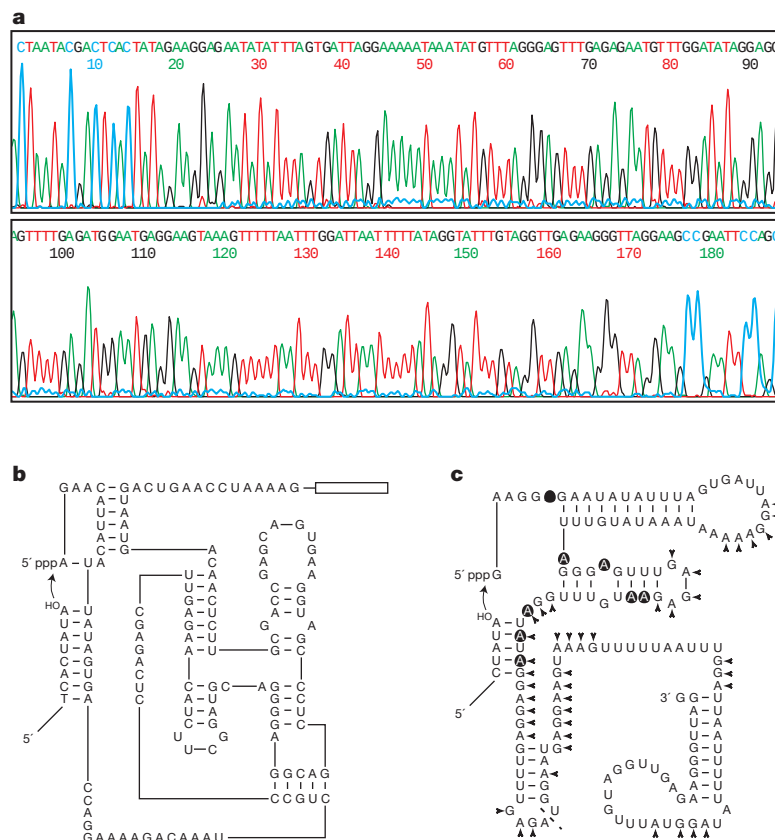


Figure 1 Composition of the final selected cytidine-free ribozyme. **a**, Sequence trace showing the lack of cytidines at nucleotide positions 19–173. Positions 2–18 correspond to the T7 promoter sequence and positions 174–188 correspond to the downstream vector sequence (pCR 2.1). Automated sequencing was carried out using an ABI model 373 DNA sequencer and was confirmed by manual sequencing of both strands (data not shown). **b**, Secondary structure of the starting ribozyme (E100) based on that of the class I

ligase¹⁰. Box indicates the primer binding site at the 3' end of the ribozyme. **c**, Secondary structure of the final selected ribozyme based on chemical modification of unpaired adenosine and guanosine residues (carat marks), carried out in the absence of substrate. Highlighted adenosine residues blocked catalytic activity when methylated at N1. Dashed line indicates the site of the largest 3'-terminal deletion that was compatible with catalytic activity.

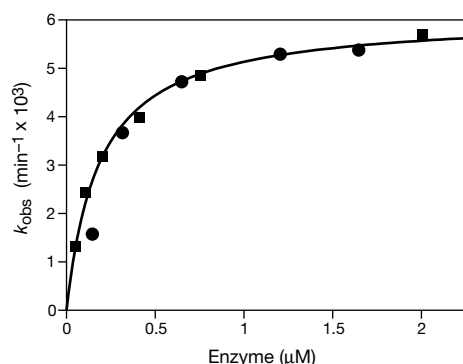


Figure 2 Catalytic activity of the final selected cytidine-free ligase. Observed rates were determined in the presence of 5 nM labelled substrate and varying concentrations of enzyme, with enzyme always in excess over substrate, as described¹¹. Reaction conditions: 25 mM MgCl₂, 50 mM KCl, 2 mM spermidine, 4 mM DTT and 50 mM EPPS, pH 8.5, at 23 °C. Data are from two independent experiments and were fit to the Michaelis–Menten equation, $v = k_{cat}[E]/(K_m + [S])$, to obtain the values for k_{cat} (0.006 min⁻¹) and apparent K_m (230 nM).

would result in the degradation of RNA. The pool of bisulphite-treated DNA molecules was converted to double-stranded DNA and then transcribed in the absence of CTP to produce a pool of cytidine-free RNA molecules. These RNAs were gel purified, reacted with the chimaeric DNA/RNA substrate, gel purified again, reverse transcribed, amplified by PCR and finally forward transcribed to complete one round of the *in vitro* selection procedure.

After eight rounds of *in vitro* selection, individuals were cloned from the population and sequenced. Most of the cloned individuals lacked cytidine, but their catalytic rate was only about 10⁻⁵ min⁻¹. One of these individuals was used to generate a pool of randomized variants, this time prepared by chemical synthesis. A population of DNA templates was synthesized using standard phosphoramidite chemistry, which introduced random mutations at a frequency of 8% per nucleotide position. Deoxyguanosine was avoided at all positions so that the corresponding RNA transcripts would not contain any cytidine.

Starting from the pool of randomized variants, eight additional rounds of *in vitro* selection were carried out. Again, individuals were cloned from the population and sequenced. One of the more active individuals, which lacked cytidine, had a catalytic rate constant (k_{cat}) of 0.006 min⁻¹ and a Michaelis constant (K_m) of 540 nM in the RNA-catalysed ligation reaction. It contained 20 mutations relative to the individual that had been randomized by chemical synthesis. All of the clones that were sequenced at this point contained a nearly identical set of mutations, suggesting that there were few distinct molecules in the chemically randomized pool with this level of catalytic activity.

Another randomized pool was prepared by chemical synthesis, this time based on the more active individual that was isolated after the sixteenth round of *in vitro* selection. Eight more rounds of selective amplification were carried out, making 24 in all, and once again individuals were cloned from the population and sequenced. These molecules remained free of cytidine, but showed no further improvement in catalytic rate. The nucleotide sequence of a typical individual isolated from the final population is shown in Fig. 1a. This ribozyme has a k_{cat} of 0.006 min⁻¹ and a K_m of 230 nM, as demonstrated by the saturation plot shown in Fig. 2. The twofold improvement in K_m compared with that of the ribozyme that had been isolated after the sixteenth round explains why the new variant emerged during the latter rounds of selection. Unlike the population that was isolated after the sixteenth round, the population

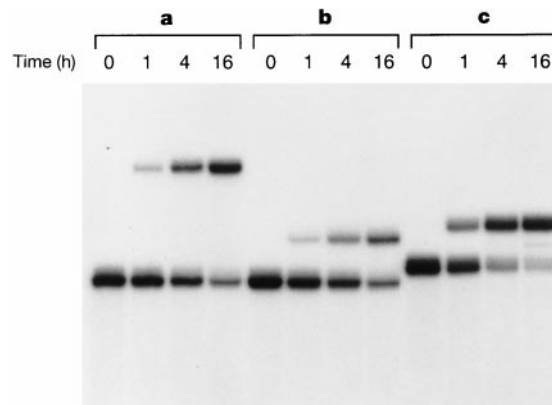


Figure 3 Autoradiogram depicting the time course of RNA-catalysed RNA ligation, using 50 nM labelled ribozyme and 10 μM substrate. The cytidine-free ribozyme derived from the E100 ligase was reacted with either the full-length substrate (**a**) (see Methods) or a shortened substrate with the sequence 5'-CGACTCACAUA-3' (RNA portion underlined) (**b**). The cytidine-free ribozyme derived from random sequence was reacted with the shortened substrate (**c**). Products were separated by electrophoresis on an 8% denaturing polyacrylamide gel.

obtained after the twenty-fourth round exhibited considerable sequence diversity. Thus, many sequence solutions exist that are closely related to the previously selected solution.

In the presence of saturating concentrations of Mg²⁺, the catalytic rate of the final selected ribozyme was 0.008 min⁻¹, with an apparent dissociation constant (K_d) for Mg²⁺ of 20 mM (data not shown). This rate is roughly 10⁵-fold greater than the uncatalysed rate of template-directed RNA ligation^{9,13}. It is significantly less, however, than the rate of the starting E100 ribozyme, which had a k_{cat} of 20 min⁻¹. Despite the lack of cytidine, the final selected ribozyme does not seem to have any difficulty folding into an active conformation; denatured ribozymes that were added to the reaction mixture reacted to a maximum extent of more than 60% (Fig. 3a, b).

To determine whether a 2',5'- or 3',5'-phosphodiester linkage was formed during ligation, the final selected ribozyme was reacted with a [5'-³²P]-labelled substrate that contained only deoxyribonucleotides, except for a single 3'-terminal ribonucleotide, and was then digested with RNase T2. RNase T2 cleaves 3',5'- but not 2',5'-phosphodiester linkages of RNA and is unable to cleave DNA. The labelled product of the reaction and subsequent digestion was one nucleotide longer than the substrate, indicating that the cytidine-free ligase forms a 2',5' linkage. This is in contrast to the starting E100 ribozyme, which forms a 3',5' linkage (data not shown).

There was no explicit pressure during *in vitro* evolution to maintain 3',5' regiospecificity or to retain any of the structural elements of the starting ribozyme. Secondary structural analysis, based on 3'-terminal deletion studies and chemical probing with dimethyl sulphate and kethoxal, revealed that the cytidine-free ligase has a different structure to that of the E100 ribozyme (Fig. 1b, c). The final selected ribozyme contains a 5'-terminal region of ~85 nucleotides that is structurally well defined and is essential for catalytic activity. There are ~70 nucleotides at the 3' end that are less constrained and can be deleted, albeit with a tenfold decrease in catalytic rate. Despite the loss of 37 cytidines between the starting and final ribozymes and the accompanying rearrangement of secondary structure, the two ribozymes have a sequence similarity of 45% and are related by descent.

A second *in vitro* evolution experiment was carried out, this time starting with a pool of 10¹⁴ random-sequence RNAs that lacked cytidine. Selection was carried out as described above, requiring the molecules to ligate a chimaeric DNA/RNA oligonucleotide to their own 5' end. Details of this experiment will be published

elsewhere, but a brief summary is given here to address the issue of how starting with a pool based on the E100 ribozyme may have influenced the outcome of evolution. After 5 rounds of selective amplification, 11 individuals were cloned from the population, analysed and sequenced. One of these individuals formed a 3',5'-phosphodiester linkage and was used to prepare a randomized pool, which was subjected to four additional rounds of *in vitro* selection. A typical ribozyme isolated from the final selected population catalysed RNA ligation in a 3',5'-specific manner, with a rate of 0.01 min⁻¹ and a maximum extent of more than 80% (Fig. 3c). It contains 158 nucleotides including 46 adenosine, 52 guanosine and 60 uridine residues. Based on terminal deletion analysis, 91 nucleotides were found to be required for catalytic activity.

The cytidine-free ribozyme that was obtained from random-sequence RNAs has roughly the same catalytic rate as the one obtained from the E100 ligase, but differs in that it generates a 3',5' linkage. Previous studies have indicated that 3',5' ligases are rare compared with 2',5' ligases in a pool of random-sequence RNAs that contain all four nucleosides¹⁰. This appears to be the case for cytidine-free RNAs as well, although both types of ligases could be obtained. Template-directed ligation between an oligonucleotide 2'- or 3'-hydroxyl and an oligonucleotide 5'-triphosphate has special significance because it involves essentially the same chemistry as that carried out by an RNA-dependent RNA polymerase. Thus, it is reasonable to contemplate the possibility of an RNA replicase ribozyme that lacks cytidine, although all four nucleosides would probably be needed to transfer genetic information.

Cytidine-free ribozymes show that evolution can cope with a very restricted set of chemical building blocks in generating macromolecules that have a complex structure and function. Several biological RNAs have very low cytidine content, such as the *dutA* messenger RNA of *Dictyostelium discoideum*¹⁴, mitochondrial RNase P RNA of *Candida glabrata*¹⁵, and various introns from dicot plants¹⁶. Cytidine-free ribozymes can be studied as model systems, with catalysis as a measure of structural integrity, to understand how naturally occurring cytidine-deficient RNAs adopt a well defined secondary and tertiary structure. How, for example, does an RNA molecule compensate for the lack of a stable G-C pair? How is a unique tertiary fold achieved when every pyrimidine (U) has the potential to pair with every purine (A and G)?

The original genetic material has been proposed to contain fewer than four different subunits, for example, a system involving only adenine and inosine¹⁷. Clearly, RNA folding and catalysis can be achieved in a three-nucleotide system. Perhaps guanosine could be eliminated as well. Biological evolution has developed an elegant solution to the problem of cytidine's chemical instability: organisms with large DNA genomes use 5'-methyl deoxyuridine (thymidine) in place of deoxyuridine and rely on the enzyme uracil-DNA glycosylase to repair deoxyuridines that have resulted from the deamination of deoxycytidine¹⁸. With or without cytidine, evolution manages to find an acceptable solution. □

Methods

Construction of starting pool

Cytidine residues were removed from the primer binding site at the 3' end of the E100 ribozyme using a mutagenic oligodeoxynucleotide to generate a new 3' terminus with the sequence 5'-...AGGUUGAGAAGGGUAGG-3'. The altered molecule was randomized at the DNA level at a degeneracy of 10% per nucleotide position by mutagenic PCR¹⁹ and then transcribed. The resulting RNAs were used as the input for 5 h of continuous *in vitro* evolution¹¹, restoring catalytic efficiency in the context of the new 3' terminus. These RNAs were used to prepare complementary DNA, which was mutagenized at both 0.7% and 10% degeneracy by mutagenic PCR^{19,20}. The two mutagenized pools were combined, resuspended in 3 M sodium bisulphite at pH 5.8 and incubated at 95 °C for 60 min. The DNA was purified by affinity chromatography on a Nensorb 20 column (NEN Life Science), treated with alkali, purified by denaturing polyacrylamide gel electrophoresis and amplified by PCR for 10 cycles.

Synthesis of randomized variants

Individuals isolated after the eighth and sixteenth rounds of selection were randomized at

a degeneracy of 8% per nucleotide position. Two synthetic oligonucleotides were prepared that contained these random mutations and had complementary sequences at their 3' ends. The oligonucleotides were gel purified, annealed and extended with *Taq* DNA polymerase. The extension efficiency was roughly 15%, resulting in 300 pmols (2×10^{14} molecules) of double-stranded DNA.

In vitro selection

DNA was transcribed in the absence of CTP using 1 pmol of DNA, 1 U μ l⁻¹ T7 RNA polymerase, 15 mM MgCl₂, 2 mM spermidine, 5 mM dithiothreitol, 50 mM Tris, pH 7.5, and 2 mM each of ATP, GTP and UTP, which were incubated at 37 °C for 1 h. Transcription was terminated by the addition of 1 U μ l⁻¹ RNase-free DNase I. The RNA was gel purified and reacted with 5 μ M substrate with the sequence 5'-CTTGACGTCAGCCTGGACT-AATACGACTCACUAUA-3' (RNA portion underlined), in a mixture containing 25 mM MgCl₂, 50 mM KCl, 2 mM spermidine, 4 mM DTT and 50 mM EPPS, pH 8.5, at 23 °C. The reacted RNA was gel purified, reverse transcribed to cDNA, ethanol precipitated and PCR amplified. During the second, fourth, sixth and eighth rounds of selection, the reacted RNA was reverse transcribed in the absence of dGTP and subsequently PCR amplified using a primer that hybridized to the last 16 nucleotides at the 3' end of the cDNA.

Analysis of secondary structure

Unlabelled, gel-purified ribozyme (50 pmol) was partially hydrolysed by incubation in 1 mM Na₂EDTA and 50 mM NaHCO₃/Na₂CO₃, pH 9.2, at 95 °C for 10 min, then ethanol precipitated and reacted with 5 pmol of [5'-³²P]-labelled substrate under the reaction conditions described above. The products were separated on an 8% denaturing polyacrylamide gel and compared with a sizing ladder to determine the smallest 3'-truncated ribozyme that could react with the labelled substrate. Chemical probing experiments were carried out under standard reaction conditions, using dimethyl sulphate and kethoxal to modify the N1 position of adenosine and N1 and N2 positions of guanosine, respectively²¹. The sites of modification were determined by primer extension analysis using reverse transcriptase and a primer that hybridized to the last 18 nucleotides at the 3' end of the ribozyme. Modification interference experiments involved partial methylation with dimethyl sulphate, followed by reaction with unlabelled substrate, separation of the reacted and unreacted ribozymes, and comparative primer extension analysis.

Received 17 May; accepted 13 September 1999.

- Gesteland, R. E., Cech, T. R. & Atkins, J. F. (eds) *The RNA World* 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999).
- Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
- Lorsch, J. R. & Szostak, J. W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36 (1994).
- Eklund, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).
- Zhang, B. & Cech, T. R. Peptidyl-transferase ribozymes: *trans* reactions, structural characterization and ribosomal RNA-like features. *Chem. Biol.* **5**, 539–553 (1998).
- Unrau, P. J. & Bartel, D. P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263 (1998).
- Levy, M. & Miller, S. L. The stability of the RNA bases: implications for the origin of life. *Proc. Natl Acad. Sci. USA* **95**, 7933–7938 (1998).
- Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
- Eklund, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
- Wright, M. C. & Joyce, G. F. Continuous *in vitro* evolution of catalytic function. *Science* **276**, 614–617 (1997).
- Hayatsu, H., Wataya, Y. & Kai, K. The addition of sodium bisulfite to uracil and to cytosine. *J. Am. Chem. Soc.* **92**, 724–726 (1970).
- Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118**, 3332–3339 (1996).
- Yoshida, H., Kumimoto, H. & Okamoto, K. *dutA* RNA functions as an untranslatable RNA in the development of *Dictyostelium discoideum*. *Nucleic Acids Res.* **22**, 41–46 (1994).
- Shu, H., Wise, C. A., Clark-Walker, G. D. & Martin, N. C. A gene required for RNase P activity in *Candida (Torulopsis) glabrata* mitochondria codes for a 227-nucleotide RNA with homology to bacterial RNase P RNA. *Mol. Cell. Biol.* **11**, 1662–1667 (1991).
- Goodall, G. J. & Filipowicz, W. The A-U rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**, 473–483 (1989).
- Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
- Vassilyev, D. G. & Morikawa, K. Precluding uracil from DNA. *Structure* **4**, 1381–1385 (1996).
- Vartanian, J., Henry, M. & Wain-Hobson, S. Hypermutagenic PCR involving all four transitions and a sizeable portion of transversions. *Nucleic Acids Res.* **24**, 2627–2631 (1996).
- Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
- Stern, S., Moazed, D. & Noller, H. F. Analysis of RNA structure using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481–489 (1988).

Acknowledgements

This work was supported by a grant from NASA.

Correspondence and requests for materials should be addressed to G.F.J. (e-mail: joyce@scripps.edu).

Out of sequence

New ways of extracting, sequencing, measuring and binding with DNA.

Spreadex Polymer NAB

Elchrom Scientific AG

An additive for acrylamide-Bis gels for electrophoresis of double-stranded DNA

Gels containing Spreadex Polymer NAB are claimed to show a resolving power 2–3-fold higher than achieved by standard acrylamide gels. A pair of DNA fragments differing by 4 bp in the 100–300-bp range is typically resolved after migrating less than 4 cm. DNA fragments in the 70–150-bp range that differ by 1 bp can be separated on gels just 8 cm long.

Reader Service No. 100

mtDNA Extractor

Wako

www.wakousa.com

Mitochondrial DNA extraction

Complementing Wako's mitochondrial DNA extraction kit for whole blood, the company has introduced a kit designed for use with cell and tissue samples. The mtDNA Extractor CT Kit protocol is simple and can be completed in about 1 hour. This kit avoids the use of hazardous reagents and can be used with fresh or frozen tissues.

Reader Service No. 101

EZ::TN Insertion Kits

Epicentre

Insert a Tn5 transposon into any DNA target

EZ::TN Insertion Kits permit random *in vitro* insertion of a natural or artificial Tn5-type transposon into a DNA. By inserting primer binding sites, EZ::TN Kits speed up sequencing of cDNA or genomic templates too large to handle with a single set of sequencing reactions. The transposon is incubated with the template DNA clone in the presence of EZ::TN Transposase for 2 hours at 37°C. Insertion clones are obtained by transforming competent *E. coli* with an aliquot of stopped reaction, and plating on selective media.

Reader Service No. 102

Gene Spec I

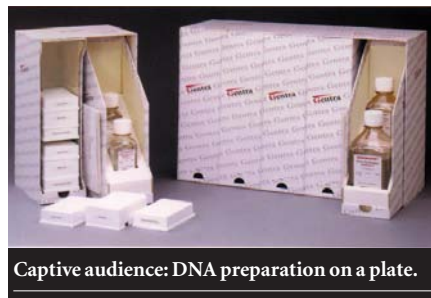
Hitachi

www.hitachi-soft.com

Quantify DNA/RNA samples down to 1 µl

With a footprint about that of an A4 sheet of paper, the Gene Spec I measures DNA/RNA samples in less than 8 seconds and displays results in a complete UV spectrum. From sample absorbance, the 260/280-nm ratio, RNA/ssDNA/dsDNA concentration, purity and protein concentration are derived instantly. Cuvettes are available in 1, 2, 5, 10, 20 and 50 µl sizes.

Reader Service No. 103



Captive audience: DNA preparation on a plate.

Capture Plate

Gentra Systems

www.gentra.com

A kit for rapid DNA sample preparation

The Capture Plate Kit eliminates pre-lysing and precipitation, enabling faster batch processing in high-throughput applications. The kit utilizes 96-well plates, manually or with automated systems, and a flexible solid-phase system with reagents that lyse cells and then capture and release DNA. This process is designed to be linked to PCR and other downstream hybridization analyses.

Reader Service No. 104

Wizard SV 96

Promega

www.euro.promega.com/uk

For rapid isolation of plasmid DNA

The Wizard SV 96 Plasmid DNA Purification System can isolate plasmid DNA from 96 samples in less than 1 hour. The plate design virtually eliminates cross-contamination between samples and allows eluted DNA to be collected in the standard 96-well plate format. The plasmid DNA can be used directly for automated fluorescent sequencing as well as other standard molecular biology techniques.

Reader Service No. 105

HB-1000 Hybridizer

UVP

www.uvp.com

A hybridization oven with precision temperature controls

The HB-1000 Hybridizer is designed to create the conditions needed for cDNA library screenings, primer synthesis and all nucleic-acid hybridization. Temperature settings are microprocessor controlled and the oven features dynamic mixing with adjustable rotation speeds. The HB-1000 has a small footprint of 445 × 406 × 381 mm.

Reader Service No. 106

Perfectprep

Eppendorf-Netheler-Hinz www.eppendorf.de

Plasmid purification for high sample yield

The Perfectprep Plasmid 96 Spin kit allows for

simultaneous isolation of plasmid DNA from up to 192 bacterial clones in 2 hours. The plasmid DNA can be used for all molecular biology applications immediately after elution without any additional precipitation, as well as for automated sequencing, transfection, blotting or cloning. All the isolation steps are performed in a 96-well plate.

Reader Service No. 107

GC-RICH PCR

Roche Molecular Biochemicals

www.biochem.boehringer-mannheim.com

Overcome templates that are GC-rich, have repeats, or exhibit high secondary structure

Roche's GC-RICH PCR System makes use of a blend of *Taq* DNA polymerase and a proof-reading enzyme, and is claimed to outperform conventional polymerases or additives. A GC-RICH resolution solution amplifies templates that are difficult or impossible to amplify by standard methods. The system is designed to cope with difficult templates up to 5 kb in length.

Reader Service No. 108

Automated sequencing

Sigma

www.sigma.sial.com

A new brochure covers automated sequencing

This updated brochure includes new products compatible with any terminator premix, with a view to cost savings. Also new are the QuickComb porous membrane loading comb for easy bench-top sample loading and the SigmaSpin size-exclusion-based post-reaction clean-up for both 96-well plates and single columns.

Reader Service No. 109



Automatic choice: sequencing.

R5319 Multichannel PMT

Hamamatsu www.hamamatsu.co.uk

A multichannel detector for DNA sequencing

Micro Titre plates (MTPs) are widely used in diagnostic instruments for DNA sequencing. The R5319 Multichannel PMT has been developed especially for detecting low light level luminescence from samples in MTPs, with single photon counting on each channel. The pitch and size of the 8 linear PMT channels match those of the MTP, making direct measurements of optical signals possible.

Reader Service No. 110

BioProbe

Enzo Diagnostics www.enzobio.com

Nonradioactive in situ hybridization systems

The newly formatted BioProbe range of *in situ* hybridization and detection systems includes

systems for labelling DNA, RNA and oligodeoxynucleotide probes with biotin, fluorescein or other labels. Enzo also supplies a range of highly sensitive detection systems (Simply-Sensitive and UltraSensitive) with a choice of fluorescent or calorimetric signal generation.

Reader Service No. 111

96EZLoad

Sooner Scientific www.soonersci.com

Eliminate carry-over contamination

Designed to simplify the loading of 96-lane DNA sequencing gels, the 96EZLoad gel loading system is a lightweight multi-channel pipette. The 96EZLoad system is available with an expandable micro-well plate to match the 96-lane upgrade spacing of the Applied Biosystems 377 DNA Sequencer. Laboratories that prefer to process samples in micro-well tube strips may also use the micro-plate

holders provided with the 96EZLoad system.

Reader Service No. 112

SteamScrubber

Labconco www.labconco.com

Clean up your act

This laboratory-grade washer cleans sequencing plates to the purity level required for DNA protein band analysis. A DNA Sequencing Plate Insert holds up to six electrophoresis plates; the washer's lower rack can accommodate two inserts at once for a total of 12 plates. Models are available that elevate water wash temperatures to 180°F (82°C).

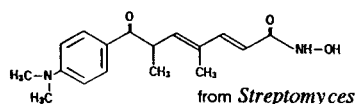
Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers.

For more details, fill in the reader service card bound in the journal.

ADVERTISEMENTS

Trichostatin A



Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call
Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

Biomolecular Synthesis oligos, synthetic peptides and polyclonal antibodies

Contact Severn Biotech Ltd
Tel: 01562 825 286
Fax: 01562 825 284
email: info@severnbiotech.com
www.severnbiotech.com

READER ENQUIRY NO. 281

WWW.LABX.COM

THE FIRST PLACE TO LOOK FOR
LAB EQUIPMENT

- Buy, Sell, & Trade with labs, dealers and manufacturers.
- 40,000 visitors per month.
- Thousands of instruments listed.
- Millions of dollars traded.

Live Online Auctions!

WWW.LABX.COM

READER ENQUIRY NO. 301

contact: info@TIB-MOLBIOL.de

READER ENQUIRY NO. 18

contact: info@TIB-MOLBIOL.de

READER ENQUIRY NO. 25

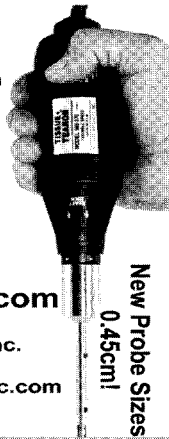
Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

**Check
Us
Out**

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com



READER ENQUIRY NO. 54

Long Read DNA Sequencing Up To 1200 bp/read

Fast Turnaround Time

Only \$50/read

MacConnell Research Corp
1-800-466-7949

READER ENQUIRY NO. 271

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray:
Phone/Fax: +44 191 222 8612
Email: JOE.GRAY@newcastle.ac.uk

DNA sequencing (ABI 377-96KL, 877 Catalyst)
also available at competitive prices

READER ENQUIRY NO. 21